Evaluation of Complex Biocatalysis in Aqueous Solution. Part I: Efforts Towards a Biophysical Perspective of the Cellulosome; Part II: Experimental Determination of Methonium Desolvation Thermodynamics

by

Jason Ryan King

Department of Chemistry
Duke University

Date:_____

Approved:

_____
Eric J. Toone, Supervisor

_____
Katherine J. Franz

_____
Dewey G. McCafferty

_____
Terrence G. Oas

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Chemistry in the Graduate School
of Duke University

2014

ABSTRACT

Evaluation of Complex Biocatalysis in Aqueous Solution. Part I: Efforts Towards a Biophysical Perspective of the Cellulosome; Part II: Experimental Determination of Methonium Desolvation Thermodynamics

by

Jason Ryan King

Department of Chemistry
Duke University

Date:_____
Approved:

_____
Eric J. Toone, Supervisor

_____
Katherine J. Franz

_____
Dewey G. McCafferty

_____
Terrence G. Oas

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the Graduate School of Duke University

2014

Abstract

The intricate interplay of biomolecules acting together, rather than alone, provides insight into the most basic of cellular functions, such as cell signaling, metabolism, defense, and, ultimately, the creation of life. Inherent in each of these processes is an evolutionary tendency towards increased efficiency by means of biolgocial *synergy*— the ability of individual elements of a system to produce a combined effect that is different and often greater than the sum of the effects of the parts. Modern biochemists are challenged to find model systems to characterize biological synergy.

We discuss the multicomponent, enzyme complex the *cellulosome* as a model system of biological synergy. Native cellulosomes comprise numerous carbohydrate-active binding proteins and enzymes designed for the efficient degradation of plant cell wall matrix polysaccharides, namely cellulose. Cellulosomes are modular enzyme complexes, comparable to enzyme "legos" that may be readily constructed into multiple geometries by synthetic design. Cellulosomal enzymes provide means to measure protein efficiency with altered complex geometry through assay of enzymatic activity as a function of geometry.

Cellulosomes are known to be highly efficient at cellulose depolymerization, and current debates on the molecular origins of this efficiency suggest two related effects

provide this efficiency: i) substrate targeting, which argues that the localization of the enzyme complex at the interface of insoluble cell wall polysaccharides facilitates substrate depolymerization; and ii) proximity effects, which describe the implicit benefit for co-localizing multiple enzymes with divergent substrate preferences on the activity of the whole complex.

Substrate targeting can be traced to the activity of a single protein, the cellulosomal scaffoldin cellulose binding module CBM3a that is thought to uniquely bind highly crystalline, insoluble cellulose. We introduce methods to develop a molecular understanding of the substrate preferences for CBM3a on soluble and insoluble cellulosic substrates. Using pivaloylysis of cellulose triacetate, we obtain multiple soluble cello-oligosaccharides with increasing degree of glucose polymerization (DP) from glucose (DP1) to cellodecaose (DP10) in high yield. Using calorimetry and centrifugal titrations, cello-oligosacharides were shown to not bind *Clostridial cellulolyticum* CMB3a. We developed AFM cantilever functionalization protocols to immobilize CBM3a and then probe the interfacial binding between CBM3a and a cellulose nanocrystal thin film using force spectroscopy. Specific binding at the interface was demonstrated in reference to a control protein that does not bind cellulose. The results indicate that i) CBM3a specifically binds nanocrystalline cellulose and ii) specific interfacial binding may be probed by force spectroscopy with the proper introduction of controls and blocking agents.

The question of enzyme proximity effects in the cellulosome must be answered by assaying the activity of cellulosomal cellulases in response to cellulosome geometry. The kinetic characterization of cellulases requires robust and reproducible assays to quantify functional cellulase content of from recombinant enzyme preparations. To facilitate the real-time routine assay of cellulase activity, we developed a custom synthesis of a fluorogenic cellulase substrate based on the cellohexaoside of Driguez and co-workers (*vide infra*). Two routes to synthesize a key thiophenyl glycoside building block were presented, with the more concise route providing the disaccharide in four steps from a commercial starting material. The disaccharide building blocks were coupled by chemical activation to yield the fully protected cellohexaoside over additional six steps. Future work will include the elaboration of this compound into an underivatized FRET-paired hexasaccharide and its subsequent use in cellulase activity assays.

This dissertation also covers an experimental system for the evaluation of methonium desolvation thermodynamics. Methonium (-N+Me3, Am) is an organic cation widely distributed in biological systems. The appearance of methonium in biological transmitters and receptors seems at odds with the large unfavorable desolvation free energy reported for tetramethylammonium (TMA+), a frequently utilized surrogate of methonium. We report an experimental system that facilitates

incremental internalization of methonium within the molecular cavity of cucurbit[7]uril (CB[7]).

Using a combination of experimental and computational studies we show that the transfer of methonium from bulk water to the CB[7] cavity is accompanied by a remarkably small desolvation enthalpy of just 0.5±0.3 kcal•mol^{-1}, a value significantly less endothermic than those values suggested from gas-phase model studies (+49.3 kcal•mol-1). More surprisingly, the incremental withdrawal of methonium surface from water produces a non- monotonic response in desolvation enthalpy. A partially desolvated state exists, in which a portion of the methonium group remains exposed to solvent. This structure incurs an increased enthalpic penalty of ~3 kcal•mol^{-1} compared to other solvation states. We attribute this observation to the pre- encapsulation de-wetting of the methonium surface. Together, our results offer a rationale for the wide biological distribution of methonium and suggest limitations to computational estimates of binding affinities based on simple parameterization of solvent-accessible surface area.

This dissertation is dedicated to my mom, aunts, uncles and grandmother who chose to sacrifice, so I could succeed; to the teachers, mentors and colleagues who have shown me how to succeed; and to my fiancée and best friend, without whom I have no reason to celebrate the success.

Table of Contents

List of Tables

List of Figures

List of Schemes

List of Equations

Acknowledgements

Foremost, I acknowledge my advisor Professor Eric Toone. My involvement in the study of cellulosomes, calorimetry, force spectroscopy, protein chemistry, and oligosaccharide synthesis originated from Eric's passion for these subjects. I thank Eric for his steadfast support, both financially and academically, of my research. With each conversation, I continue to glean a chemist's intuition from Eric that will undoubtedly prove invaluable throughout the lifetime of my career in science.

I acknowledge my dissertation committee members. As a course instructor and research consultant, Professor Terrance Oas fostered my excitement for scientific research and challenged me to address difficult questions with the confidence afforded from an appreciation for experimental assumptions and physical constants of molecular systems. Professor Katherine Franz has been a valuable faculty member in my efforts towards the PhD. During my preliminary and original research proposal examinations Professor Franz pushed me to focus my thoughts and energy so that my writing and presentation styles reflect my research aims clearly and concisely. Professor Dewey McCafferty has been a constant presence in my time at Duke. As an instructor, Dewey inspired me to learn and add to the vibrant living story of mechanistic enzymology. I view Dewey as a surrogate advisor, allowing me open access to the many resources in his laboratory and providing constant emotional support during my progress to the PhD.

The work presented in this dissertation is a result of numerous contributions from collaborators and co-workers. Yi Wang is a wonderful friend and collaborator who originally developed the idea of incremental internalization in host-guest systems in order to study functional group desolvation energies and performed most of the computational analyses presented in chapter 4. Carleen Bowers is an amazing scientist and colleague from whom I learned all of the surface chemistry and force spectroscopy detailed in chapter 2. I acknowledge Professor David Beratan and Pan Wu for their contributions to the work presented in chapter 4. Current and former Toone Lab members Brian Watts, Andrea Luteran, Briana Vogen, Erin Wilfong, David Gooden, Allison Schmitt, David Carlson, Alex Shestopalov, Jim Parise, and Matt Walters provided invaluable experimental know-how and emotional support. I thank Small Molecule Synthesis Facility staff members Ramesh Gopalaswamy, and Xiaofei Liang for expert advice in organic synthesis. I thank current and former Mcafferty Lab members Amanda Hoertz, Maria Bednar, Sunhee Hwang, Katherine Clancy, Jonathan Burg, and Kenneth Maksimchuk for technical support of biochemical techniques and instrumentation. I thank former undergraduate students Charles Saadeh and Justin de Biasio who worked tirelessly in the synthesis of cellodextrins.

I thank the Craig Lab, Zauscher lab, and Liu Lab at Duke for necessary instrument time and support. I acknowledge Professor Carol Fierke and the Fierke Lab of the Univeristy of Michigan for DNA design and experimental details. I thank

Professors Ed Bayer from the Weizmann Institute in Rehovot, Israel and Henri-Pierre Fierobe of the University of Marseilles, France for the generous gift of Clostridial cellulases. Duke university staff members Anthony Ribeiro, Ron Venters, George Dubay, Todd Woerner, and Mark Walters provided expert advice and time. I acknowledge the Duke University Graduate School, Department of Chemistry and the Burroughs-Welcome Foundation for financial support.

1. The cellulosome: structure and function from genetic profiling to structural biology.

1.1 The cellulosome in the era of post-reductionist biochemistry

In his 1998 recollection on the "golden era for understanding enzyme mechanisms," Professor Gordon G. Hammes described mechanistic enzymology—or the study of the myriad molecular processes that confound the chemistry of enzyme-catalyzed reactions—as a vibrant area of study dominated by meticulous research into minutest details of an enzymatic transformation.[1] His opinions were largely influenced by the advent of highly sensitive tools in the 1990's that allowed scientists to measure biochemical heat flows as small as 10 nanocalories per second, protein dynamics in the picosecond regime, and biomolecular phenomena with single molecule resolution. One might say Professor Hammes' account marks the pinnacle of the *Reductionist Era* of biochemistry. Almost overnight, however, the "dissect and build" mantra of the post-reductionist movement in contemporary chemical biology that came to the fore at the turn of the century quickly and drastically altered the way we study the chemistry of biology.

Gierasch and Gershenson beautifully illustrated the new direction of biochemistry with a somewhat whimsical perspective entitled "Post-reductionist protein science, or putting Humpty Dumpty back together again," which appeared in *Nat. Chem. Biol.* in November of 2009.[2] This account correlates the puzzled attempts of "all the king's men" to

reassemble Humpty Dumpty to the efforts of modern biochemists to elucidate enzyme mechanisms out of the context of the cell with the aim to determine the cellular function of the enzyme and its associated pathways. To equate this story to my own education, my advisor once related reductionist biochemistry to studying all of the pieces of a car to learn to drive. Bruce Albert writes that "indeed, the entire cell can be viewed as a factory that contains an elaborate network of interlocking assembly lines, each of which is composed of a set of large protein machines."[3]

Post-reductionist biochemistry is the study of biomolecules using cellular systems and in-vitro model systems that represent the cellular environment. Cellular and multi-protein model systems are larger than traditional single protein experiments, but can still be genetically or biochemically modified to impose controls and limits, such as the suppression of undesired side reactions via gene silencing. The goal of post-reductionist biochemistry is not to discount or remove detailed studies of individual biomolecules for in-depth mechanistic study, but rather to explore these molecules in a larger context with the aim of illustrating the new phenomena that arise from the complex interplay of molecular processes.

Herein, we describe an experimental system tailored to monitor a ubiquitous characteristic of biology, *synergy*—the ability of individual elements of a system to produce a combined effect that is different and often greater than the sum of the effects

of the parts—, as a function of increased chemical complexity. Biological synergy is a characteristic of a group of molecules whose combined activity is evidently different and often greater than the sum of the contributions of the parts (e.g. enzyme velocity in multi-enzyme systems or avidity in extracellular carbohydrate binding proteins). A molecular description of synergy is almost entirely absent from the current state of knowledge. Here, we attempt to model biological synergy with aims to elucidate the chemical and kinetic mechanism of the Clostridial enzyme system known as the *cellulosome*. Made up of individually active protein components that collectively function with enhanced activity in comparison to the sum of the parts, the cellulosome provides a unique opportunity to build complexity into an enzyme system and simultaneously quantify synergy as is pertains to this complexity.

1.1.1 The cellulosome: discovery and initial characterization

In the early 1980s, Israeli scientists Edward Bayer and Raphael Lamed reported the discovery of a polyp-shaped protrusion on the exterior cell membrane of *Clostridia thermocellum*, an anaerobic, thermophilic bacterium isolated from sewage for to its ability to efficiently degrade cellulose, or the $\beta(1 \rightarrow 4)$-polymer of D-(+)-glucose.[4-8] Using electron microscopy (Figure 1b and c; scale bars ~ 100 nm) and immunoelectrophoresis Lamed, Bayer, and others correlated the presence of these polyps with the ability to bind Avicel (microcrystalline cellulose).[4]

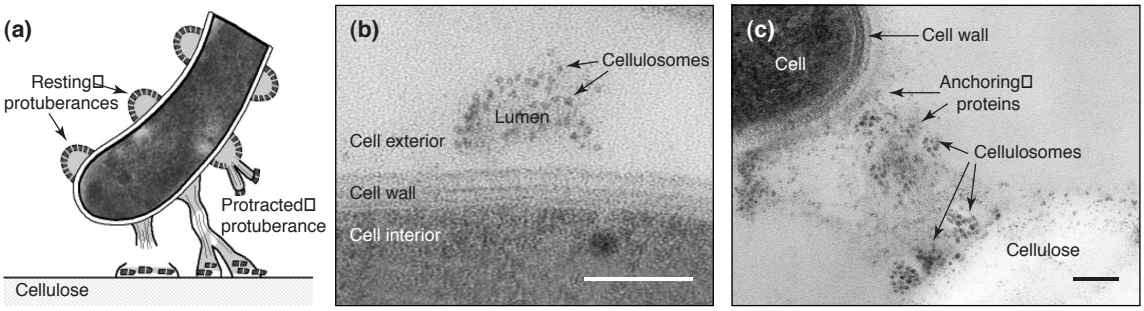


Figure 1: Microscopic images of cellulosome protrusions on *C. thermocellum* (adapted from Shoham et al.)[6] a) diagram of cell bound to cellulose; b & c) TEM images of cellulosomes.

Bayer's data, combined with previous reports from Petre et al.[7] and Demain[9] that the cellulase activity of *C. thermocellum* was largely extracellular, gave rise to the notion of the cellulosome as an extracellular cellulase complex. Bayer and colleagues coined the term "cellulosome" to mirror other classifications of megadalton cellular protein machinery and organelles such as the canonic nucleosome, ribosome, and liposomes.[6, 10] The structure of the cellulosome was determined by sedimentation, electrophoresis, gel permeation, and electron microscopy as a large, megadalton protein-based molecule.[4, 5, 11] Surprisingly, once subjected to SDS-PAGE, the single 2 MDa protein was found to comprise fourteen lower molecular weight proteins ranging in size from 50 – 200 kDa.[5]

Extensive analysis of the cellulosome structure using electron microscopy identified cellulosomes as large as 3.5 MDa containing up to 50 protein subunits in various strains of *C. thermocellum*.[12] Ljungdahl and co-workers also reported so-called "polycellulosomes" comprised of multiple (6-15) cellulosomes with tight and loose geometries on the order of 50-80 MDa and 60 nm in diameter.[12] Ljungdahl's micrographs

offered the first experimental evidence for key structural features of the cellulosome. High-resolution images of 25 nm cellulosome particles displayed chain-like fibers with globular proteins attached in an ordered array along the length of the fiber (Figure 2a). The images revealed the cellulosome to be a fibrous complex of globular proteins, now known as cellulase enzymes, along a supporting scaffold protein, termed the *scaffoldin* protein. The resolution of these original images would not be improved until 2011 when the dissect and build strategy of Bayer and Smith combined X-ray crystallography and cryogenic electron microscopy (Cryo-EM; silver density maps in Figure 2b and c) and provided the most resolved representation of the cellulosome structure to date.

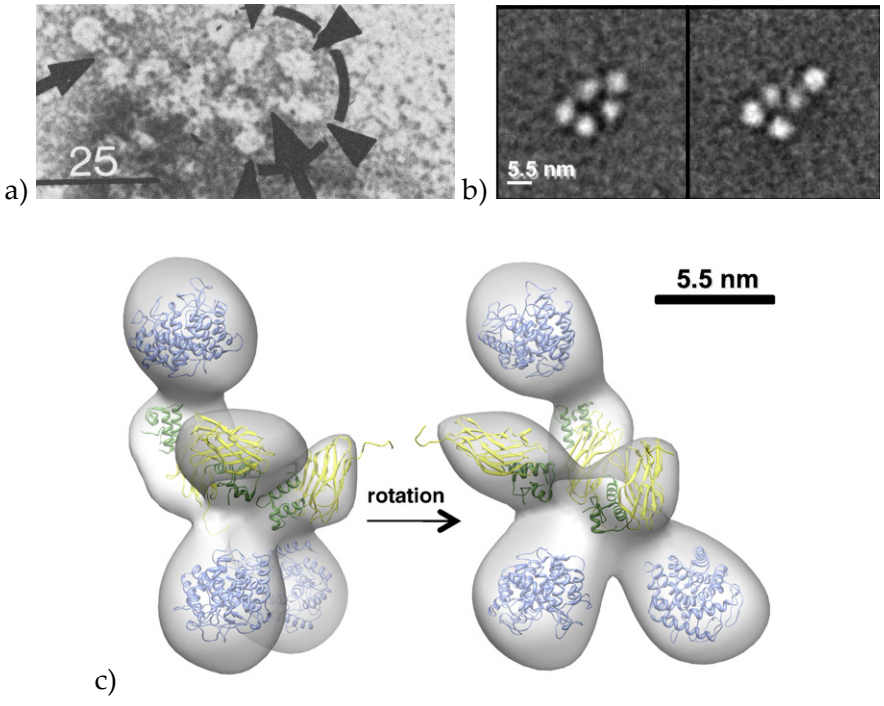


Figure 2: Cellulosome viewed by electron microscopy (a) and joint CryoEM (b & c) and X-ray crystallography (c) (adapted from Mayer et al. and Garcia-Alvarez et al.).[12, 13]

1.1.2 In search of the cellulosome paradigm for enzyme complexation

The *C. thermocellum* cellulosome archetype has been observed in numerous anaerobic bacteria including thermophillic Clostridia (*C. josui*),[14] mesophilic Clostridia (*C. cellulolvorans*,[15-22] *C. cellulolyticum*,[23-29] *C. cellobioparum*[30], *C. papyrosolvens*,[31-33] *C. acetobutylicum*),[34-36] *Acetivibrio cellulolyticus*,[37] *Bacteriodes cellulosovlens*,[38, 39] *Butyrivibrio fibrisolvens*,[40] and Rumminococci (*R. flavefaciens*,[41-43] *R. albus*,[44, 45] and *R. succinogenes*).[43] Each cellulosome employs enzyme complexation onto a structural scaffoldin protein—often denoted in the original literature as a cellulose-integrating protein (e.g. CipA from *C. thermocellum* and CipC from *C. cellulolyticum*).[14, 27] Genetic sequence analysis of the CipA protein in the *C. thermocellum* cellulosome indicated the scaffoldin contained a series of 9 repeating domains, seven of which shared 98-100% homology.[46] Later, a similar set of domain repeats was deduced for the scaffoldin protein of *C. cellulovorans*, further supporting the notion of evolutionary connections between the complexed enzymes among Clostridia.[15] These domains would later be called *cohesins* and shown to be important in the binding of cellulosomal cellulases to Cip proteins.[47]

Genetic profiling also revealed that cellulosomal enzymes were distinct from other fungal cellulases due to the presence of a 70 amino acid protein domain that proved ubiquitous among cellulosomal cellulases and scaffoldins. This 70 residue domain contained dual 22 amino acid repeats and a calcium-binding F-hand segment of the canonical EF-hand motif in calcium binding proteins. Furthermore, many of the

cellulases (particularly the exo-cellulases that degrade the cellulose polymer from the free chain ends) lacked cellulose-binding domains (CBD), and instead contained only the 70 residue domain, while other endoglucanases contained both a CBD and the 70 residue domain. The 70 residue segment, denoted as the *dockerin* domain, was a hallmark of cellulosomal proteins and quickly became a key indicator in the identification of new cellulsomes and cellulosomal proteins by phylogenetic analyses.[6]

Two significant findings illustrated the chemical simplicity behind the repetition and structural complexity in the megadalton cellulosomes. The first, and possibly the most impactful, finding was the discovery that the scaffoldin subunits (cohesins) specifically bound to the dockerin-containing enzyme CelD of *C. thermocellum*.[48] Though Aubert and co-workers originally described only the binding of a degradation product of the scaffoldin protein to the dockerin-borne CelD enzyme, a tremendous amount of work has since detailed the chemical, structural, and thermodynamic details of the cohesion-dockerin interaction.[13, 27, 37, 38, 49-56] At once, the observations of repeating patterns of cohesion domains along scaffoldin proteins and the ubiquitous inclusion of the 10 kDa dockerin subunit into cellulosomal enzymes provided a clear indication of cellulsomal assembly. Futhermore, as the cohesion repeats along the scaffoldin were nearly identical,[46] the assembly of dockerin-borne enzymes along the scaffoldin appeared as a stochastic process, a notion soon supported by the reports of Bayer and others.[57-59]

Soon,[48] the initial characterization of the cellulose binding properties of the *C. thermocellum* CipA scaffoldin and, later,[60] the identification of the type 3a cellulose binding module (CBM3a) as responsible for anchoring the entire cellulosome to crystalline cellulose prior to cellulolysis was reported. The fine-tuning of cellulose recognition by CBMs as a function of cellulose structure is described below, but the evolutionary advantages that the scaffoldin-borne CBM3a provides cellulosomal organsims are clearly apparent. Unlike fungal exo-cellulases that each contain a CBM3a or similar subunit for association to cellulose, the cellulosome provides a single anchoring CBM3a domain far removed from the catalytic domains. The result of this geometry is an apparent advantage for the processive movement of exo-cellulases along the surface of the cellulose. The two paradigmatic attributes of the *C. thermocellum* cellulosome (the cohesion-dockerin orchestrated cellulosome assembly and a single scaffoldin-borne anchoring CBM) are preserved across the various structural types of different cellulsomal systems.

1.1.2.1 The cellulosomes of *C. thermocellum* and mesophilic clostridia

The cellulosome architecture of *C. thermocellum*, elucidated through numerous genetic and biochemical studies from the collaboration of Bayer, Lamed, and Béguin, is displayed in Figure 3.[28, 30, 46-48, 51, 53-55, 60-67] The results indicated an intricate network of enzyme-bound primary scaffoldins, each of which contains an orthogonal dockerin domain (type II) that binds the type II cohesion repeats of one of three cell-bound

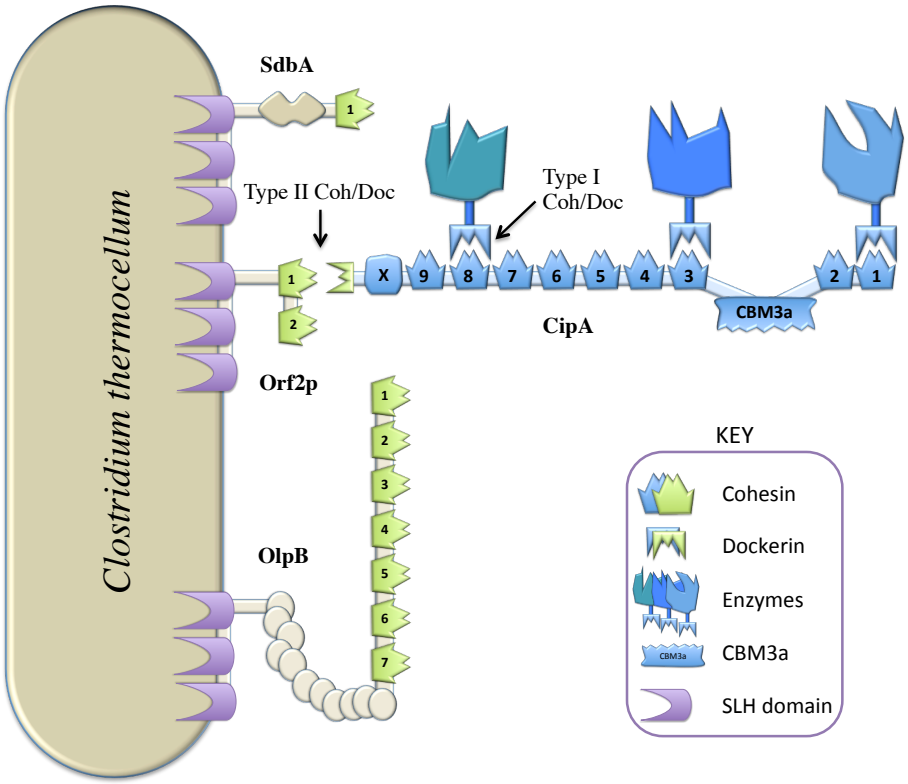


Figure 3: Clostridium thermocellum cellulosome (adapted from Bayer, 2008)[47]

anchoring scaffoldins (SdbA, Orf2p, and OlpB). The enzyme-borne dockerin was classified as type I and the cognate cohesins found in the primary CipA proteins used for enzyme immobilization were also designated as type I. The anchoring scaffoldins incorporated ten type II dockerins across three anchoring scaffoldins (SdbA: 1; Orf2p: 2; and OlpB: 7) allowing incorporation of up to ninety enzymes, given that each CipA protein contains nine type I cohesins. Unlike the CipA protein, which contains a CBM3a domain, the anchoring scaffoldins contain only type II cohesins and S-layer homology

(SLH) domains that anchor the cellulosome to the cell membrane. Recent reports also suggest two additional CipA-like scaffoldins lacking surface-anchoring SLH repeats may be expressed.[68] These proteins have not been isolated, but genomic analysis suggests they contain multiple type-II dockerin domains that may aid in formation of oligomerizing CipA scaffoldins. Indeed such oligomerization was reported by Bayer et al., and, interestingly, oligomerization was disrupted by the *in vitro* introduction of a recombinant type II dockerin-infused beta-glucosidase into the cellulosome.[69] Thus, the complexity of the multiple scaffoldin domains includes a level of organization in which the primary scaffoldins (CipA) bear enzymes that degrade cellulose and CBM3a domains that target insoluble, crystalline cellulose matrix, while, secondary anchoring scaffoldins dock the CipA•enzyme complexes to the cell membrane where, presumably, the organism consumes the products of cellulolysis.

Comparison of the *C. thermocellum* cellulosome *C. josui*[*] and mesophilic Clostridia revealed a divergent cellulosome structure that lacked the type II cohesion/ dockerin domains and anchoring scaffoldins. The cellulosomes of *C. cellulolytiucm*, *C. cellulovorans*, and *C. josui* are not cell associated, and each contains a single enzyme-binding scaffoldin domain with surprisingly similar modular structures (Table 1). The striking contrast between the cellulosome structure of *C. thermocellum* and the remaining Clostridia

---

[*] *C. josui* has been the subject of some debate as to its thermophilic or mesophilic character. It is currently described as a moderate thermophile.[70. Sukhumavasi, J., Ohmiya, K., Shimizu, S., and Ueno, K. (1988) Clostridium josui sp. nov., a cellulolytic, moderate thermophilic species from Thai compost, *International Journal of Systematic and Evolutionary Microbiology 38*, 179-182.]

Table 1: Clostdrial scaffoldin proteins

Organism	Scaffoldin	Modular Structure	MW (kDa)	Ref.
C. thermocellum	CipA	2(Coh1)-CBM3a-7(Coh1)-X-Doc2	197	46
C. cellulovorans	CbpA	CBM3a-X-2(Coh1)-X-6(Coh1)-2X-Coh1	189	15
C. cellulolyticum	CipC	CBM3a-X-7(Coh1)-X-Coh1	157	71
C. josui	CipA	CBM3a-X-6(Coh1)	120	14

sparked a debate about the evolutionary development of the cellulosome. To resolve this debate, extensive phylogenetic analyses of the cohesion and dockerin domains of cellulosome-containing organisms were performed (Figure 4).[28]

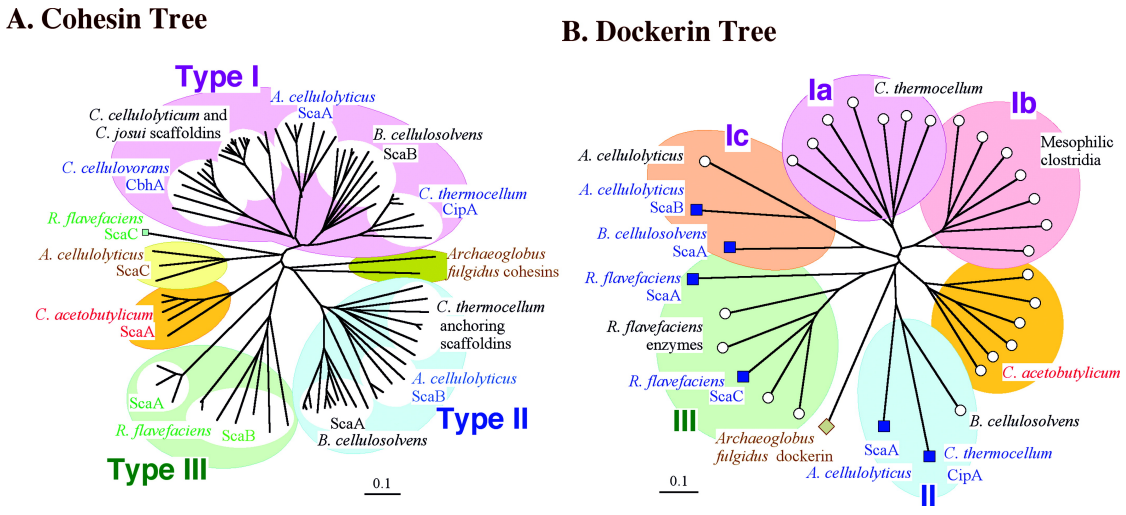


Figure 4: Phylogenetic tree of cellulosomal cohesins and dockerins[28†]

The results confirmed a close relationship among the divergent types of type I, II, and III (*vide infra*) cohesins and dockerins. Not surprisingly, *C. thermocellum* was found to be

† Annual Review of Microbiology by Annual Reviews. Reproduced with permission of Annual Reviews in the format Republish in a thesis/dissertation via Copyright Clearance Center.

evolutionarily removed from the other mesophilic Clostridia, which appeared to share a common ancestor.

Genomic mining of cellulosome-encoding Clostridial DNA revealed a rather surprising gene cluster of scaffoldin and enzyme open reading frames. Again, the enzyme-rich gene clusters of mesohilic Clostridia differed strikingly from the scaffoldin gene cluster of *C. thermocellum*.[28, 47, 72] Aside from highly homologous enzyme-linked gene clusters for the cellulosomes of mesophilic Clostridia, the species were genetically diverse. It has been postulated that the similarities in cellulosomes came not from divergent evolution of a common ancestor but from horizontal gene transfer across Clostridial communities of similar origin. Indeed, *C. cellulolyticum*, *C. cellulovorans*, and *C. josui* were all isolated from agricultural waste.[18] It is interesting to note that *C. acetobutylicum*, a well-known solvent-producing clostridium, contains an enzyme-linked cellulosome gene cluster, but does not produce a cellulosome.[34, 47, 73]

1.1.2.2 The complex cellulosomes of *Acetevibrio cellulolyticus*, *Bacteriodes cellulosolvens*, and *Ruminoccus flavefaciens*

Table 2: Complex cellulosome scaffoldins

Organism	Scaffoldin	Modular Structure	Ref.
C. thermocellum	CipA	2(Coh_1)-CBM_{3a}-7(Coh_1)-X-Doc_2	[46]
	SdbA	Coh_2-(linker)-3(SLH)	
	OlpB	7(Coh_2)-(linker)-3(SLH)	
	Orf2p	2(Coh_2)-3(SLH)	
A. cellulolyticus	ScaA	GH9-3(Coh_1)-CBM_{3a}-4(Coh_1)-X-Doc_2	[37, 47]
	ScaB	4(Coh_2)-Doc_3	
	ScaC	3(Coh_3)-3(SLH)	
	ScaD	2(Coh_2)-Coh_1-3(SLH)	

B. cellulosolvens	ScaA	5(Coh_2)-CMB_{3a}-6(Coh_2)-Doc_1	*38, 39, 47*
	ScaB	10(Coh_1)-X-3(SLH)	
R. flavefaciens	ScaA	Coh_X-2(Coh_1)-Doc_2	*41, 47, 74, 75*
	ScaB	4(Coh_1)-4(Coh_2)-Doc_X	
	ScaC	Coh_3-Doc_1	
	ScaE	Coh_{3e}	
	CttA	2(CBM_{ND})-Doc_X	

*ScaE is covalently linked to the cell membrane of *R. flavefaciens* in a sortase-mediated process. ND: the family of CBM in CttA has yet to be determined.

Genome mining of several cellulose-utilizing organsims revealed *A. cellulolyticus* along with other non-Clostridia (below) as cellulosome-producing organisms.[37] The *A cellulolyticus* cellulosome presented several "improvements" on the *C. thermocellum* paradigm (Table 2). An intriguing difference was the incorporation of an additional adapter scaffoldin (ScaB) that connects the primary (ScaA) and anchor (ScaC and ScaD) scaffoldins, which, in turn, show similarity to CipA, OlpB, and Orf2p of the *C. thermocellum* cellulosome. Like the Clostridial cellulosomes, the *A. cellulolyticus* cellulosome presents only the CBM3a protein needed to bind crystalline cellulose within the primary scaffoldin, ScaA. Unlike the clostridial Cip proteins, ScaA also contains an endo-cellulase domain (glycohydrolase family 9, GH9). With the additional scaffoldin ScaB comes an added level of complexity, as the adapter scaffoldin multiplicatively increases the enzyme loading capacity of the cell-bound scaffoldins. With the ability to bind orthogonal dockerin domains with ScaD, the *A. cellulolyticus* cellulosome is capable of incorporating 193 different enzymes.

The *B. cellulosolvens* cellulosome, discovered by Lamed and co-workers,[38, 39] also diverges structurally from the *C. thermocellum* archetype with the implementation of two massive scaffoldin domains: the primary ScaA scaffoldin that contains eleven type II cohesins and a CBM3a module, and an anchoring scaffoldin (ScaB) that presents 10 type I cohesins in conjunction with SLH cell-anchoring domains. As with the ScaA protein, the *B. cellulosolvens* enzymes also contain type II dockerin domains. When fully loaded, these two *B. cellulosolvens* scaffoldins contain as many as 110 enzymes.

By far, the most complex known cellulosome is produced by the anaerobic rumen bacterium *Ruminoccocus flavefaciens*. After the initial discovery of dockerin-contianing proteins in *R. flavefaciens* by Flint and co-workers, an international collaboration detailed its cellulosome components.[41, 42, 75, 76] To date five scaffoldin proteins—ScaA, ScaB, ScaC, ScaE and CttA—have been characterized. Unlike the previously reported cellulosomes, no clear primary scaffoldin protein is evident. Rather, the scaffoldin protein CttA, which binds directly to the cell anchoring scaffoldin ScaE contains two putative carbohydrate binding modules of unknown classification. It is presumed that the series of cohesion-containing scaffoldins (Table 2; ScaA, ScaB, ScaC, and ScaE) organizes dockerin-bearing enzymes in a similar manner to the *A. cellulolyticus* cellulosome. Lacking the CBM3a, this series of scaffoldins is likely targeted to cellulose via CttA and their co-localization on the exterior of the cell.[47, 74]

1.1.3 Structural plasticity and development of designer cellulosomes

1.1.3.1 Structural insights into the cohesin-dockerin interaction

Despite the variety and number of cellulosomal proteins characterized to date, the cellulosome can be simplified to a series of calcium-dependent cohesion-dockerin interactions with divergent specificities. These "Lego-like" domains have been the subject of extensive study since the mid 1990's. The cohesion, dockerin (usually enzyme-bound), and complexes of the two domains have all been characterized by X-ray crystallography. The bound structure depicts a 1:1 complex with an apparent dual binding mode (i.e. a C2 axis of symmetry exists such that the dockerin could presumably bind a second cohesion domain via a single displacement mechanism on the opposite face of the bound cohesion.[52] Characterization of the *C. thermocellum* type I cohesin-dockerin interaction via isothermal titration calorimetry shows the binding to be entropically driven at 298 K with a binding free energy of -12.9 kcal•mol^{-1} and ΔC_P = -305 cal•mol^{-1}K^{-1},[77] parameters consistent with the proposed hydrophobic binding. Close examination of the cohesion-dockerin interaction revealed cohesin specificity determinants on the surface of the dockerin domain. After experimentally confirming an intra-species preference for the cohesin-dockerin interaction using dockerin and cohesin dyads from *C. thermocellum* and *C. cellulolyticum*, Bayer and co-workers used protein sequence alignment to identify key residues responsible for this specificity. The so-called specificity determinants were found to be the 10th and 11th amino acids of each 22 amino acid repeat in the dockerin domain. All of the known type I dockerins from *C.*

thermocellum exhibited two polar, hydroxyl-containing residues (SS or ST) at positions 10 and 11 while the *C. cellulolyticum* dockerins displayed non-polar amino acids at these positions (AL, AI, or AV).[58] The pivatol role of these residues was confirmed through site-directed mutagenesis and X-ray crystallography in the early 2000's.[49, 78-80] The initial designation of type I cohesin-dockerin interactions as "species specific" was later found to apply only to thermophilic versus mesophilic Clostridia, as dockerins from different mesophilic Clostridia contain the same residues at positions 10 and 11. Evaluation of the *C. cellulovorans* and *C. josui* dockerin specificity determinants suggests they should bind to *C. cellulolyticm* and *C. thermocellum* cohesins, respectively [Ed Bayer; personal communication].

1.1.3.2 Designer miniature cellulosomes

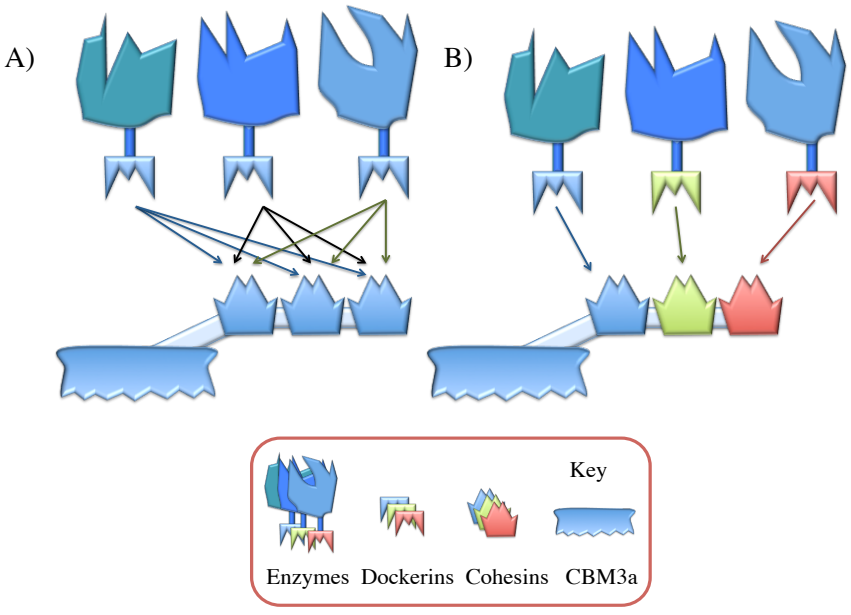


Figure 5: Designer miniature cellulosomes

With an understanding of the types and specificities found in cohesin-dockerin pairs, Bélaïch and coworkers began a collaboration with the Bayer group to engineer designer miniature cellulosomes with a goal of creating biological catalysts suitable for the industrial depolymerization of lignocellulose. These "minicellulosomes" comprised two or three cellulases from *C. thermocellum* or *C. cellulolyticum*, a family IIIa carbohydrate binding module (CMB3a), and a single bi- or tri-modular chimeric scaffoldin domain containing the corresponding number of orthogonal cohesin domains from *C. thermocellum* or *C. cellulolyticum* (Figure 5b). Unlike wild type megadalton cellulosomes, these minicellulosomes more closely resemble the protein lego architecture gleaned from the parent proteins.

By varying the type and number of cohesins on the scaffoldin and the location of the CBM3a, Bélaïch and co-workers were able to demonstrate that incorporation of even a single scaffoldin domain afforded the complexed enzymes an apparent advantage in lignocellulose hydrolysis over their non-complexed counterparts. Dockerin-containing enzymes were mixed with avicel and the reducing sugar equivalents produced during cellulolysis were assayed using dinitrosalicylic acid. To ensure efficient cellulolysis, two cooperatively functioning cellulases (a family 9 endoglucanase, CelF, and a family 48 exoglucanase, CelS) were used. Cellulolysis was evaluated for non-complexed enzymes and compared to complexed analogues. To demonstrate the role of the CBM3a, as opposed to the scaffoldin, on cellulolysis, each enzyme was bound to a monovalent

scaffoldin containing a CBM3a and then assayed for cellulase activity. Though the targeting effect of the CBM3a-bound cellulases notably improved the cellulolysis, an apparent synergistic improvement (4.1-fold increase) in the cellulase activity was observed for the miniature scaffoldin.

In the years following initial studies with designer miniature cellulosomes, an extensive body of work by the former students of Bélaïch, Fierobe and Tardif has explored numerous miniature cellulosome geometries with scaffodin chimeras, incorporating various combinations of Clostridial and fungal (non-cellulosomal) enzymes with engineered dockerin domains. An exquisite account of the functional consequences of enzyme proximity and substrate targeting on cellulolysis is arguably the most impactful contribtuion from these studies. Current debate regarding the physical origin of cellulosome activity centers around questions of enzyme proximity vs. substrate targeting. Several studies have attempted to address this issue using small, well-defined miniature cellulosomes, but interpretation of the results remains contentious (*vide infra*).

1.2 The biological degradation of cellulose

Cellulose, the β-(1 → 4)-linked polymer of D-(+)-glucose, is the most abundant polymer on Earth. Cellulose is found in nature in numerous forms that vary considerably in crystallinity and the presence of other biopolymers. Largely due to the recalcitrance of cellulose in nature, cellulose-utilizing organisms, namely anaerobic

bacteria of the *Clostridia* and *Bacteroides* genera and aerobic fungi such as *Trichoderma reesei*, have evolved to produce a multitude of cellulose-degrading enzymes that catalyze the hydrolysis of internal (endocellulases) or external (exocellulases) glycosidic bonds.[81, 82] Exocellulases (also commonly known as cellobiohydrolases, cellobiases, and exoglucanases) are further categorized according their preference for non-reducing (cellobiohydroase I) or reducing (cellobiohydrolase II,) termini of the polymer chain.[82] The ubiquity of cellulose in plants and plant-based products and the potential of cellulose as a renewable carbon source for liquid fuels and bio-based polymers provides motivation to understand cellulase activity in both isolation and in the context of cellulolytic organisms.

1.2.1 The structure and bio-availability of cellulose

The monomeric building block of cellulose is D-(+)-glucopyranose. The degree of polymerization (DP) refers to the number of glucose units in a chain and ranges from 30 - 15,000. The carbons of the glucose ring are referenced as C1-C6, beginning with the anomeric carbon in the pyranose form. Alternating glucose units are rotated 180° with respect to the orthogonal axis of the pyranose ring, and thus a convenient functional building block of cellulose is the dimer of glucose, cellobiose (Figure 6a). The most processive cellulose degrading enzymes (cellobiohydrolases) actually remove consecutive units of cellobiose, rather than glucose, from the cellulose chain.

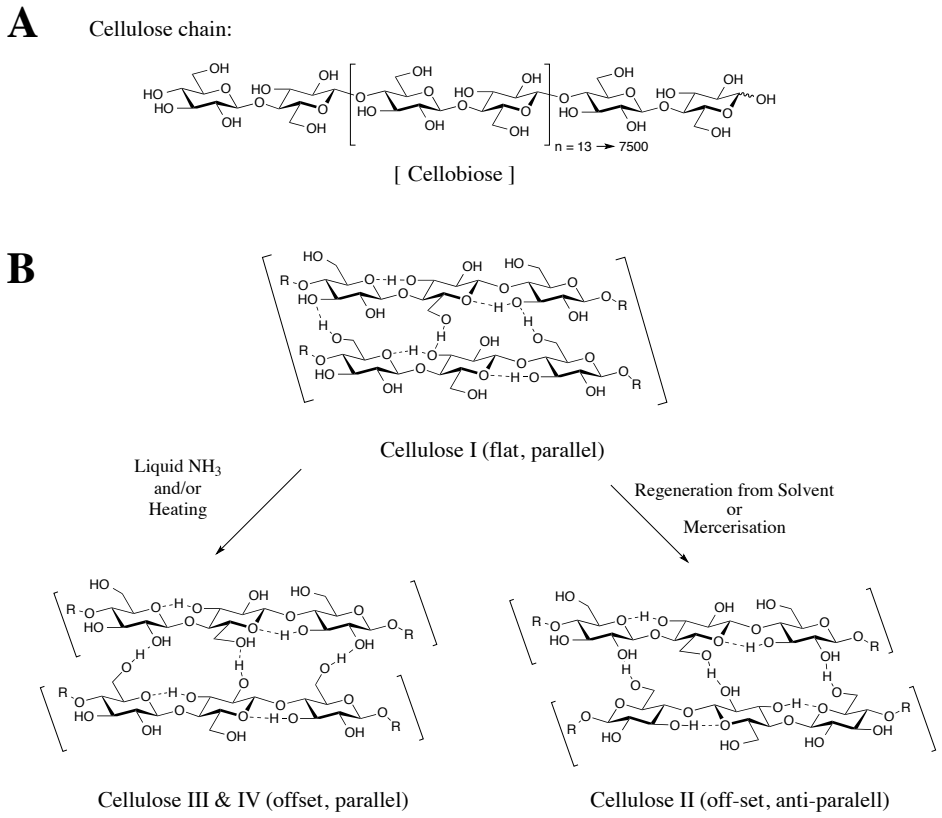


Figure 6: Structure of Cellulose

Figure 6 depicts the structure of a single cellulose chain and the four polymorphs of cellulose (I, II, III, and IV). Cellulose I, which is further categorized into the two allomorphs I_α and I_β, is the most abundant form of cellulose found primarily in microfibril cores of the plant cell well. Industrially, cellulose I has been modified to cellulose II (regenerated cellulose) by dissolution in a solvent such as N-methyl morpholine oxide (NMMO) followed by precipitation with water. Alternatively, cellulose II can be produced by alkaline swelling of cellulose I in a process known as

mercerisation. Cellulose III is produced via ammonia treatment of cellulose I, and cellulose IV is formed through thermal rearrangement of cellulose III.

The structures of cellulose polymorphs have been extensively studied by microscopy, X-ray diffraction, and X-ray chrystallography.[83-86] At a molecular level, cellulose I is characterized by a flat sheet of parallel cellulose chains with inter-chain hydrogen bonds between C3 and C6 hydroxyls (C3O---H—OC6) (Figure 6). Stabilizing intra-chain C3-OH pyranose oxygen interactions limit flexibility of the C3 hydroxyl moiety, leading to a tightly packed structure that is largely hydrophobic in nature. Regenerated cellulose (cellulose II), which is often used as a model substrate for cellulase studies in the form of cellulose filter paper, exhibits a less compact crystal structure than cellulose I. In a notable difference from native cellulose, where interacting chains are parallel, the cellulose II chains are anti-parallel and slightly juxtaposed along the Z-axis (perpendicular to the XY plane of the cellulose chain). The chains interact through stabilizing C2 and C6 hydroxyl hydrogen bonds that flow in the Z-axis direction. Cellulose III and IV have inter-chain hydrogen bonding networks and chain positions similar to cellulose II, although the chains, like native cellulose, are parallel. The minor difference in cellulose III and IV arises from the definition of the unit cell, which is hexagonal in the case of cellulose III and orthogonal in cellulose IV.

Cellulose is a principle component of primary and secondary plant cell walls, which accounts for its ubiquity in the biosphere. Cellulose biosynthesis begins with the

production of linear cellulose chains by the plant cellulose synthase (CESA) proteins.[87] Tens of chains then spontaneously bundle into elementary fibrils, presumably due to H-bonding and solvent exclusion. Elementary fibrils are 2.5 – 4 nm in diameter and can grow to many microns in length.[86, 87] The elementary fibrils further bundle to form cellulose microfibrils. A very high aspect ratio is characteristic of cellulose fibrils, allowing for continuous wrapping of single fibrils around the circumference of a plant cell.[86, 87] Upon entering the primary cell wall matrix, the fibril most likely associates physically with other matrix polysaccharides that are deposited from the golgi apparatus.

The matrix polysaccharides include pectins—complex polysaccharides comprised of rhamnogalacturonan I and II, homogalacturonan, xylogalacturonan, and arabinan—and hemicelluloses—cellulose-binding polyfuranosides such as xyloglucan and arabinoxylan.[82, 87] Matrix polysaccharides form a hydrogel-like polymer network arising from physical association and chemical and enzymatic cross-linking of the pectin components. Though the actual mechanism of cell wall formation is unknown, the simultaneous secretion and cross-linking of cellulose microfibrils with matrix polysaccharides likely explains the extensive network of plant cell wall polysaccharides (Figure 7).[87]

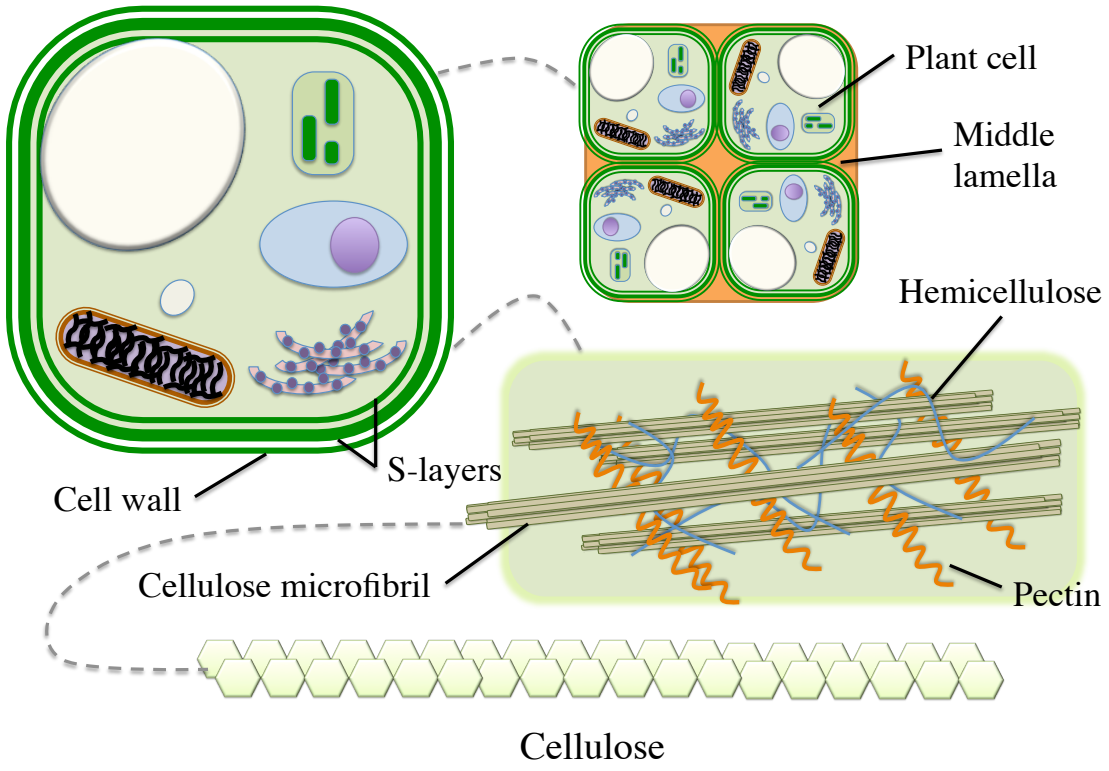


Figure 7: Cellulose structure in plant cell walls (adapted from Cosgrove, 2005)[87]

In addition to cellulose, hemicellulose, and pectin, woody plants incorporate a waxy polyphenolic structure known as lignin into the primary cell wall matrix. The primary cell wall is surrounded by a layer of lignin called the middle lamella. The lignin content correlates well with the waxy character of the plant. An interior secondary plant cell wall, comprising three layers (S1, S2, and S3), is surrounded by the primary cell wall and the middle lamella. For woody plants, the majority of cellulose is located in the S-layers of the secondary cell wall (Figure 7).[86]

1.2.2 Cellulose degradation by polysaccharide hydrolases

Cellulose manipulation, both in the biosynthesis and break-down of cellulose, requires an intricate interplay of several polysaccharide hydrolases—enzymes that catalyze the addition of water across a C-X bond, where X = O, N, resulting in some degree of polysaccharide depolymerization.[82, 88, 89] Cellulose-consuming microorganisms, including anaerobic and aerobic bacteria and fungi, commonly deplore polysaccharide hydrolases for the depolymerization of plant cell matrix polysaccharides.[72] Plants also utilize cellulases and xylanases in cell wall biosynthesis.[87]

Polysaccharide hydrolases make up the majority of glycoside hydrolases (GHs). To date, thousands of GHs have been identified. To catalog these enzymes, an international, open access, online database known as the Carbohydrate-active enzyme (CAZy) database was proposed and created by Henrissat.[90] The database organizes GHs by amino acid sequence similarity and is currently managed by the team of Cuotinho and Henrissat (www.cazy.org).[91] A large group of hydrolases, commonly called "hemicellulases," includes xylanases, mannases, and arabinases. Hemicellulases expose pure cellulose microfibrils from disordered strands of hemicellulose rendering the cellulose susceptible to hydrolysis.[18, 72, 82] Cellulophiles typically produce an arsenal of hemicellulases to degrade and manipulate hemicelluloses that vary vastly in xylose, mannose, and arabinose content, degree of sugar branching, and pattern of glycosidic bond ($1 \rightarrow 3$ vs. $1 \rightarrow 4$ connections). Another set of enzymes known as "pectinases,"

including polygalacturonases and pectin lyases, disrupts the matrix polysaccharides that encapsulate hemicellulose and cellulose microfibrils. Even so exposed, the pure cellulose microfibril poses challenging depolymerization problem. Three classes of cellulase are required for efficient cellulolysis: endocellulase (EG), cellobiohydrolase I and/or II (CBH I and CBH II), and β-glucosidase (BG).[82]

Endocellulases (EC3.2.1.4), designated EG for β(1→4)-endoglucanases, are the most diverse class of cellulases. EGs catalyze the hydrolysis of β(1→4)-glucan linkages in crystalline and amorphous cellulose, beta-glucan containing β(1→3)-linkages, cereal barley glucan, functionalized cellulose materials such as cellulose resins and water-soluble carboxymethyl cellulose, and cello-oligosaccharides of DP ≥ 4.[91] EG activity has been reported in fourteen families of glycoside hydrolases (GHs): GH5-GH10, GH12, GH26, GH44, GH45, GH48, GH51, GH74, and GH124.[91] The most common cellulosomal EGs are from family GH5‡ and GH9 (Table 3), though a re-interpretation of the GH5 superfamily classification into 53 subfamilies has recently suggested that cellulosomal GH5-family EGs may have evolved to certain level of substrate promiscuity—particulary those in the GH5_2 subfamily.[92] In general, EGs are multi-domain proteins that contain a catalytic GH domain ligated to a carbohydrate binding module (CBM). As described above, cellulosomal enzymes typically present a dockerin domain in place of

‡ Recently, the GH5 superfamily has been parsed into 53 sub-families to accommodate the large diversity of observed enzyme activities in this family. EGs have been reassigned to GH5 subfamilies GH5_1-5, GH5_22, and GH5_25-26, GH5_37, and GH5_39. All CBH I enzymes from the GH5 family were re-classified as GH5_1. (Aspeborg, 2012)

the CBM, though it is not uncommon to observe additional CBM domains with cellulosomal EGs.[18] In the case of GH9, the enzyme bears a CBM3b or CBM3c domain capable of binding single chains of polysaccharides.[89] EGs are necessary for the efficient degradation of crystalline cellulose, opening accessible sites for the CBHs that degrade from the cellulose chain end. This so-called "exo-endo" synergism is a common attribute of all cellulase enzyme systems.

CBH I (EC3.2.1.91) is a designation of exo-acting cellulases known to degrade single cellulose chains via the (typically processive) removal of cellobiose units from the non-reducing end of the polysaccharide (i.e. from the terminal C4′-OH). CBH I enzymes are typically modular, containing a catalytic GH domain and a CBM domain connected by a flexible linker peptide. CBH Is are active on microcrystalline cellulose (avicel), but exhibit much greater activity on amorphous cellulose and/or EG-treated cellulose. The CBH preference for end-chain cellulose sites versus the EG preference for internal cellulose chains is the basis for the well-document exo-endo synergism among cellulases.[72, 81, 82, 93-96] CBH I activity has been reported in GH families 5_1, 6 and 9.[91] Interestingly, all of these families also display EG activity, suggesting that the hydrophobic cluster analysis (HCA) classification of GH activity based solely on amino acid sequence is not sufficient to provide unique cellulase classification.[82, 90] Typically, cellulosomal GH5_1 and GH9 enzymes only present EG activity (Table 3), although GH5_1 (CelO) from *C. thermocellum* does not.[97] Although the HCA method appears to

correctly group enzymes with similar three-dimensional structures, it is likely that, given ubiquity of cellulases in biology, Nature has evolved multiple protein folds to enable cellulose degradation in various environments. Complementary to the HCA method, CBH I enzymes are also classified mechanistically via their activities on para-nitrophenyl glycosides of lactose, cellobiose, and small cello-oligosaccharides.

CBH II enzymes (EC3.2.1.176) are exo-cellulases that hydrolyze single cellulose chains from the reducing terminus. The polysaccharide substrate preferences for CBH II enzymes are similar to CBH I enzymes, favoring amorphous cellulose over avicel. A notable difference from CBH II enzymes is the apparent lack of activity against choromophoric aglycones of glucose, cellobiose, cellotriose, and cellotetraose glycosides. A molecular rationale for this behavior was presented that suggested that unlike CBH I enzymes, CBH II enzymes hydrolyze $\beta(1 \rightarrow 4)$-glycosidic linkages of cellobiose units, rendering para-nitrophenylcellotrioside as the simplest substrate for CBH II. Because this scission does not release a chromophoric aglycone, it is not useful for colorimetric assays of cellulase activity.[97] Like CBH I, CBH II enzymes are modular, containing linked CBM and GH domains. CBH II activity has been found for GH families 7, 9, and 48.[91] GH48-containing processive exocellulases, such as CelS in *C. thermocellum*, generally constitute the only CBH II activity in cellulosomal systems, although the majority of cellulolysis in cellulosomes is attributed to GH48-containing enzymes.[82, 89]

Table 3: Cellulosomal cellulases and structures

Organism	Enzyme	Type*	Modular Structure**	Ref
C. thermocellum	CelA	EG	GH8-Doc1	*6, 89*
	CbhA	CBH I	CBM4-Ig-GH9-2(Fn3)-CBM3b-Doc1	
	CelB	EG	GH5-Doc1	
	CelD	EG	Ig-GH9-Doc1	
	CelE	EG/CE	GH5-Doc1-CE2	
	CelF	EG	GH9-CBM3c-Doc1	
	CelG	EG	GH5-Doc1	
	CelH	CBH I	GH26-GH5-CBM9-Doc1	
	CelJ	EG	CBM30-Ig-GH9-GH44-Doc1-UN	
	CelK	CBH I	CMB4-Ig-GH9-Doc1	
	CelN	EG	GH9-CBM3c-Doc1	
	CelO	CBH I	CBM3b-GH5-Doc1	
	CelQ	EG	GH9-CBM3c-Doc1	
	CelR	EG	GH9-CBM3c-Doc1	
	CelS	CBH II	GH48-Doc1	
	CelT	EG	GH9-Doc1	
	XynA	XYN	GH11-CBM4-Doc1-NodB	
	XynB	XYN	GH11-CBM4-Doc1	
	XynC	XYN	CBM22-GH10-Doc1	
	XynD	XYN	CBM22-GH10-Doc1	
	XynY	XYN/ CE	CBM22-GH10-CBM22-Doc1-CE1	
	LicB	LIC	GH16-Doc1	
	ChiA	CHI	GH18-Doc1	
	ManA	MAN	CBM-GH26-Doc1	
	XghA	XG	GH74-CBM2-Doc1	
	XynZ	XYN/ CE	CE1-CBM6-Doc1-GH10	
C. cellulovorans	ExgS	CBH II	GH48-Doc1	*18*
	EngH	EG	GH9-CBM3c-Doc1	
	EngK	EG	CBM4-Ig-GH9-Doc1	
	EngL	EG	GH9-Doc1	
	ManA	MAN	Doc1-GH5	
	EngM	EG	CBM4-Ig-GH9-Doc1	
	EngE	EG	3(SLH)-GH5-X-Doc1	
	EngY	EG	CBM2-GH9-Doc1	
	EngB	EG	GH5-Doc1	

	PelA	PL	X-CBD2-PL9-Doc1	
	XynA	XYN	GH11-Doc1-CE4	
C. acetobutylicum	CelF	CBH II	GH48-Doc1	[18]
	CelA	EG	GH5-Doc1	
	CelH	EG	GH9-CBM3c-Doc1	
	EngA	EG	GH44-Doc1	
	CelG	EG	GH9-CBM3c-Doc1	
	CelL	EG	GH9-Doc1	
	ManA	MAN	GH5-Doc1	
	CAC0919	SIA	GH74-Doc1	
	CelE	EG	CBM3-Ig-GH9-Doc1	
	CAC3469	EG	3(SLH)-GH5-X-Doc1	
C. cellulolyticum	CelF	CBH II	GH48-Doc1	[18]
	CelC	EG	GH8-Doc1	
	CelG	EG	GH9-CBM3c-Doc1	
	CelE	EG	CBM4-Ig-GH9-Doc1	
	CelH	EG	GH9-CBM3c-Doc1	
	CelJ	EG	GH9-CBM3c-Doc1	
	ManK	MAN	Doc1-GH5	
	CelD	EG	GH5-Doc1	
	CelA	EG	GH5-Doc1	
	CelM	EG	GH9-Doc1	
C. josui	CelD	CBH II	GH48-Doc1	[18]
	CelB	EG	GH8-Doc1	
	CelE	EG	GH9-CBM3c-Doc1	
	AgaA	a-Gal	GH27-Doc1	

*EG: β(1→4)-Endoglucanse; CBH: Cellobiohydrolase; CE: Carbohydrate Esterase; CHI: Chitanase; LIC: Lichenase; MAN: Mannase; XYN: Xylanase; PL: Pectate Lyase; SIA: Sialidase; XG: Xyloglucanase; a-Gal: Alpha-galactanase.
**GH: Glycoside hydrolase; CBM: Carbohydrate binding module; Doc: Dockerin; Ig: Immunoglobulin-like domain; Fn: Fibronectin-like domain; SLH: S-layer homology domain; X: Hydrophilic domain; UN: domain of unknown function.

The principle product of EG, CBH I and CBH II cellulolysis is cellobiose. Cellobiose, in turn, is a potent inhibitor of EG, CBH I and CBH II. As such, efficient fungal cellulase systems incoporporate beta-glucosidases (BG; EC3.2.1.21) that hydrolyze cellobiose to glucose. BG is a monomeric enzyme with a GH domain from GH

families 1, 3, 5_9, 5_12, 9, 30, and 116. Exogenous addition of *Aspergilus niger* BG to cellulosomal systems and the addition of an engineered dockerin-fused BG to *C. thermocellum* has been reported to enhance cellulolysis, presumably by depleting cellobiose inhibition.[69, 98]

1.3 Cellulases and the interfacial enzyme model

Biological cellulolysis is a complex process requiring the combined activity of myriad hydrolase enzymes. The preceding sections were devoted to describing the classes of enzymes required for the efficient deconstruction of plant cell walls. Here we describe the chemical and kinetic characterizations of cellulases, which comprise the majority of polysaccharide hydrolases in the cellulosome. By way of context, we provide a brief historical account of the adaptation of traditional kinetic characterizations of solution-phase enzymes at steady state using the Michaelis-Menten equations to interfacial enzymes via Gelb's interfacial kinetic model.

1.3.1 Chemical characterization of cellulases

As with many other glycoside hydrolases, cellulases incorporate two aspartate and/or glutamate residues in the enzyme active site. The two residues exist in an orthogonal protonation state and promote acid/base catalysis in the addition of water to cleave the glycosidic bond of cellulose. The molecular mechanism of water substitution at the newly-formed lactol carbon may proceed in two ways, resulting in either retention or inversion of the stereochemical configuration at C1.

Inverting glycosidases are found among endoglucanases such as *C. cellulolyticum* CelB or *C. thermocellum* CelD, and exocellulases such as *C. thermocellum* CelK and CbhA. More generally, inverting cellulases are better grouped by common GH families: 6, 8, 9, 45, 48, 74, and 124.[91] The chemical mechanism of inverting cellulases has been detailed in *C. thermocellum* CelD and involves a single, concerted reaction step (Figure 8).[99, 100] A protonated glutamic acid residue positioned above the sugar serves to activate the C1 glycosyl ether of the bound sugar chain. Simultaneously, a basic aspartate deprotonates a water nucleophile which, in turn, displaces the equatorial glycosyl ether via an S_N2 axial attack at C1.

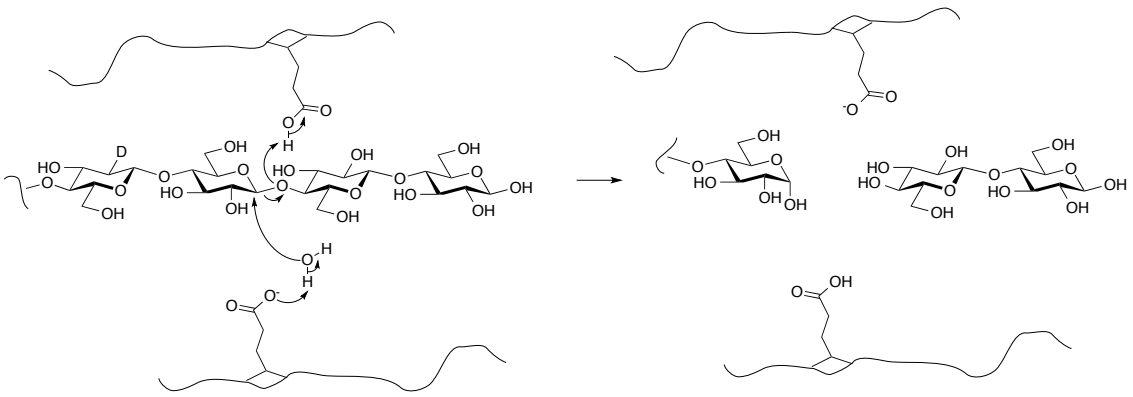


Figure 8: Inverting glycosidase mechanism

Retaining glycosidases are also found among endoglucanases and exocellulases in families GH 5, 7, 10, 12, 26, 44, and 51.[91] Because the GH5 superfamily contains the largest group of cellulass, most cellulases execute hydrolysis with retention of configuration. The retaining mechanism of β-glucanase is well demonstrated by the

paradigmatic *Cellulomonas fimi* Cex, a 1,4-β-glycosidase that shows activity against xylan and glucan.[101] The retaining β-glucanase mechanism (Figure 9) involves sequential inverting displacements and begins with the proton-activation of the C1 glycosyl ether oxygen by a protonated aspartic or glutamic acid positioned above the sugar in the enzyme active site. S_N1 cleavage of the anomeric C-O bond leads to an oxocarbonium-stabilized carbocation which is, in turn, attacked by a glutamate or aspartate base from below the plane of the sugar. This axial attack results in a glycosyl enzyme intermediate with an inverted configuration at the anomeric carbon. Back-side attack at the anomeric carbon of the glycosyl enzyme intermediate by a water nucleophile reinstates the original configuration at the anomeric carbon, and the ionized lactol is de-protonated by the glutamate or aspartate from above the sugar to regenerate the enzyme acid catalyst.

The nature of the glycosyl enzyme intermediate in retaining β-glucanases has been a topic of considerable debate.[82, 101, 102] It is well accepted that the retaining glycosidase lysozyme presents a tight ion pair between the ionized aspartate base positioned below the stable pyranosyl oxocarbonium intermediate. β-glucanases were presumed to follow a similar mechanism, though Koshland had proposed a double-displacement mechanism involving a discrete glycosyl enzyme intermediate.[103] Finally, Withers and co-workers kinetically trapped a glycosyl-enzyme intermediate with a 2-deoxy-2-fluoro-sugar substrate. Due to the deactivating fluoro substituent, subsequent

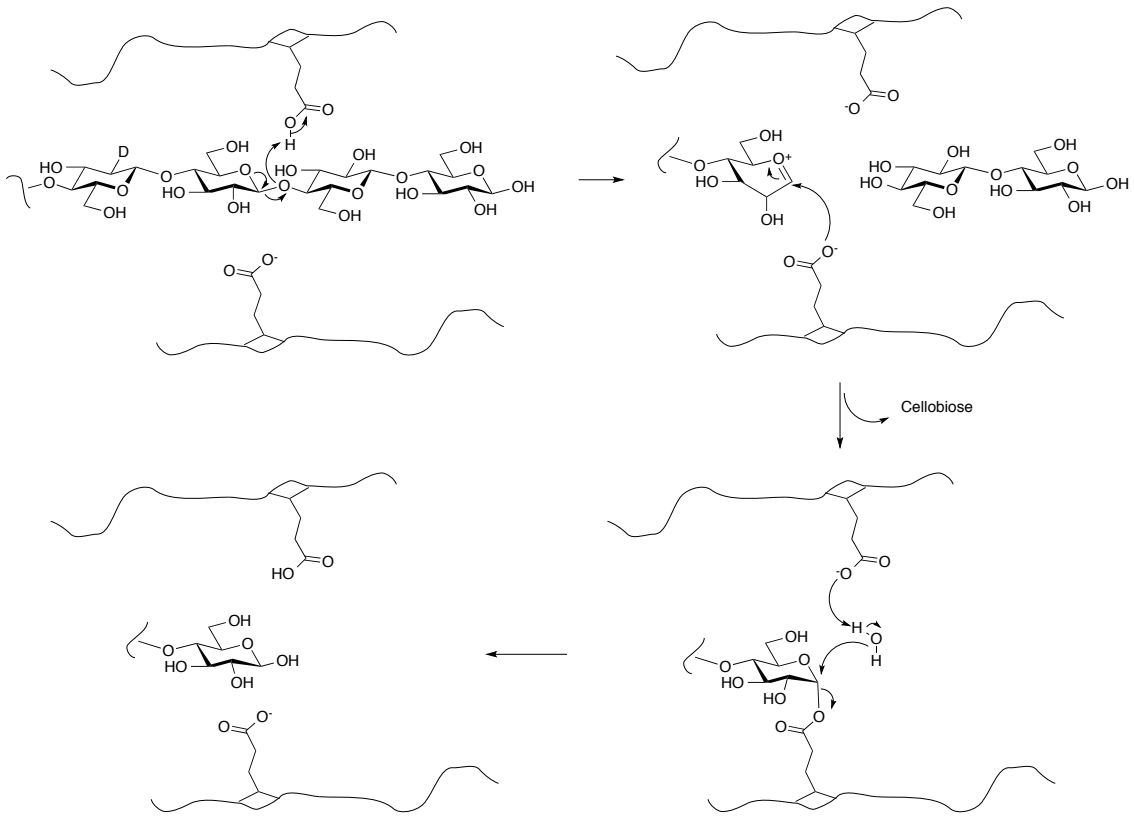


Figure 9: Retaining CBH II mechanism

hydrolysis of the intermediate was slow enough to allow capture of the glycosyl-enzyme complex in a crystal structure of *Cellulomonas fimi* Cex catalytic domain.[101]

1.3.2 The Gelb model of interfacial enzyme kinetics

Many deductions regarding the mechanism of glycosidase action rely on kinetic evaluation. Standard (Michaelis-Menten) enzyme kinetics are derived for soluble enzymes acting on soluble substrates. Cellulose is remarkably insoluble in aqueous solution, and new kinetic models are required for consideration of interfacial enzyme action on insoluble substrates. Gelb and co-workers[104] have developed such an approach

for lipase activity, and those models are applicable to cellulose action. Where appropriate, modifications to the model to accommodate cellulase-specific kinetics are presented.

Gelb and co-workers originally developed the interfacial enzyme model to describe the kinetics of lipases at the solution-vessicle interface.[104] Lipases catalyze the hydrolysis of ester linkages on micelles formed by the spontaneous self-assembly of fatty acid triglycerides and phospho-lipids. The Gelb model of interfacial enzyme activity is shown schematically in Figure 10.

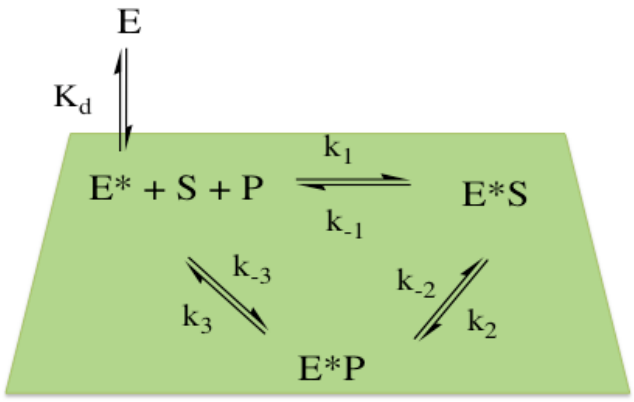


Figure 10: Gelb model of interfacial enzymes[104, 105]

At the outset, the conventional definition of substrate concentration ([S]) and product concentration ([P]) in mol•L^{-1} is modified to accommodate the two-dimensional nature of interfacial enzyme substrates. As such, the concept of *surface concentration* of substrate sites ([S*]), rather than solution concentration ([S]), is used. The mole fraction of substrate surface sites (X_S) is a convenient form of the amount of substrate sites that yields the mass balance relationship

$$X_S + X_P = 1$$

Equation 1

where X_P describe the mole fractions of catalytically incompetent surface sites and product sites (i.e. hydrolyzed glucan sites on a cellulose surface). For heterogeneous cellulosic substrates, multiple catalytically non-competent sites are present that inherently reduce the surface density of substrate sites. To account for the unused area of the surface, the surface density (ρ_S) measure of substrate concentration is taken as effective concentration:

$$\rho_S = \frac{X_S}{a_A X_A + a_S X_S} = \frac{X_S}{a_A(1 + \varphi X_S)}$$

Equation 2

$$\varphi = \frac{a_S}{a_A} - 1$$

Equation 3

where a_S or a_A is the surface area contributed the cellulase substrate (S) or general cellulose (A) surface sites, and φ is an area correction factor defined by Equation 3. The term $1/(1+\varphi X_S)$ can be considered as an activity coefficient that approaches unity for the "dilute" cellulose surfaces where $a_A >> a_S$.

To express of [S] and [P] as X_S and X_P, the conversion factor N_S^0 has been introduced, N_S^0 is the initial concentration of enzyme-accessible surface sites on the

substrate surface given in units of moles surface site per liter of bulk solvent. In the context of cellulases, N_S^0 can be experimentally determined by titrating cellulose with free enzyme and plotting moles $E*$ (immobilized enzyme that is measurably removed from the bulk solution) vs. moles E (enzyme remaining in solution during the titration). The asymptotic value of $E_{\max}^*$ required to saturate substrate sites yields N_S^0 via Equation 4:

$$E_{\max}^* = \frac{m_S}{M_S} N_S^0$$

Equation 4

where m_S denotes the mass of substrate in grams and M_S is the substrate concentration in units of grams per liter of solvent. From this relationship, it follows that

$$[P] = N_S^0 X_p$$

Equation 5.

The depletion of surface sites during the enzyme-binding titration is negligible for most cellulases, given the disparity in the kinetics of cellulose hydrolysis relative to cellulose binding. Alternatively, N_S^0 can be determined by mutagenesis of the catalytic aspartate or glutamate residues to glutamine in order to silence the catalytic activity of the enzyme; this mutation has no more than minimal effect on substrate binding.[106]

As with solution enzymes, interfacial velocity (v) is defined as the rate of product formation over time:

$$v = \frac{d[P]}{dt} = \frac{d[N_S^0 X_P]}{dt} \approx N_S^0 \frac{d[X_P]}{dt}$$

Equation 6.

v can be correlated to the mole fraction of substrate sites using the following simplifying assumptions: (i) N_S^0 does not itself vary with time; this condition may not hold for multi-cellulase systems; (ii) the carbohydrate binding module binds cellulose prior to the formation of the enzyme-substrate complex and hydrolysis of the substrate, and the CBM remains bound during the catalytic turnover of substrate to product;[107] and (iii) the hydrolysis reaction is irreversible in bulk water (i.e. $k_{-2} = 0$). It then follows that:

$$\frac{v}{N_S^0[E_T]} \approx \frac{\frac{k_{catS}}{K_{mS}} X_S}{1 + \frac{X_S}{K_{mS}} + \frac{X_P}{K_P} + \frac{K_d}{[M]}}$$

Equation 7

$$k_{catS} = k_2$$

Equation 8

$$K_{mS} = \frac{k_{-1}k_3 + k_2 k_3}{k_1(k_2 + k_3)}$$

Equation 9

where $[E_T]$ is the total enzyme concentration in mol•L^{-1}, k_{catS} is the turnover number for formation of product from the E^*S complex in units of s^{-1} or min^{-1}, K_{mS} (unitless) denotes the Michaelis constant with respect to E^* and S association, K_P is a unitless

dissociation constant for product and bound enzyme sites, K_d is the dissociation constant for free enzyme and the interface in mol•L^{-1}, and $[M]$ is the active concentration of interface binding sites in mol•L^{-1}. A notable distinction between $[M]$ and $[S]$ implies that not all interfacial binding events may to catalytically competent substrate associations, i.e. the two binding events are treated independently. This condition is certainly applicable to cellulases where the catalytic GH domain is often functionally separate from the CBM.

Under initial conditions, where X_P is negligible and $X_S \rightarrow 1$, the familiar form of the Michaelis-Menten equation is obtained:

$$\frac{v_0}{[E_T]} = \frac{k_{catS} N_S^0}{N_S^0[1 + K_{mS}(1 + \frac{K_d}{[M]})]} = \frac{k_{cat}^{app}[S_0^*]}{[S_0^*] + K_m^{app}}$$

Equation 10.

The condensed form of Equation 10 contains an apparent turnover number (k_{cat}^{app}), in units of s^{-1} or min^{-1}, and an apparent Michaelis constant (K_m^{app}) in units of mol•L^{-1}, which serve as experimental curve-fitting parameters in a plot of enzyme velocity against initial substrate concentration ($[S_0^*]$). Note that N_S^0 has been represented as $[S_0^*]$ to produce the canonical form of the Michaelis-Menten equation. A mathematical description of these apparent values in terms of k_{catS}, K_{mS}, and K_d are provided by Equation 11 and Equation 12:

$$k_{cat}^{app} = \frac{k_{catS}}{1 + K_{mS}}$$

Equation 11

$$K_m^{app} = \frac{K_d K_{mS}}{1 + K_{mS}}$$

Equation 12

The Gelb model implies that interfacial enzyme systems do not respond to global changes in substrate concentration by reaction dilution, and thus can be more accurately described by separately quantifying the interfacial adsorption/ binding of the enzyme to the substrate interface and the surface density of substrate sites.

1.3.3 Cellulase kinetics[88]

The study of cellulase kinetics covers a broad field, ranging from qualitative measurments of released soluble sugar from the insoluble cellulose matrix to quantitative, real-time measurements of cellulases activity at the single-molecule level. Cellulase assays are crucial to the study of cellulase kinetics, and these assays are discussed in detail in Chapter 3. Here, we introduce the mathematical modeling of cellulase-catalyzed cellulolysis.

Implicit in the Gelb model of interfacial enzymes is a need to quantify the innate interfacial adsorption of enzyme to substrate (K_d) and the enzyme capacity for substrate (K_{mS}). These factors are the foundation for cellulase kinetic models, though these models generally elaborate on the Langmuir isotherm expression to arrive at a similar endpoint

(*vide infra*). The goal of cellulase kinetics is to experimentally determine a rate constant, k, with respect to the measured cellulose hydrolysis rate, r_c, and concentration of the cellulase-cellulose complex, $[CE]$, via the equation:

$$r_c = k[CE]$$

Equation 13.

As with most enzyme models, $[CE]$ is, itself, not generally measurable,[§108, 109] and must be derived from a relationship between initial cellulase ($[E_T]$) and cellulose ($[S_T]$) concentrations. As noted above, a distinction must be made between the total substrate used (S_T) and the free cellulase-available surface sites (C). Active cellulase-cellulose binding equilibria are thought to follow the partition of the free cellulase (E) between C and CE as follows:

$$C + E \xleftrightarrow{K_d} CE$$

Equation 14

$$K_d = \frac{[E][C]}{[CE]}$$

Equation 15

$$\frac{d[CE]}{dt} = k_f[E][C] - k_r[CE]$$

Equation 16

---

§ A promising approach to monitor the CE complex in real-time using ultra-fast atomic force microscopy has been reported by Uchihashi et al (Uchihashi, 2012).

where K_d is the thermodynamic dissociation constant for formation of the CE complex, k_f is the 2nd order rate constant for CE formation, k_r is the 1st order rate constant for the dissociation, and $[CE]$ is defined using the enzyme partition function as a difference of free and total cellulase:

$$[CE] = [E_T] - [E]$$

Equation 17.

The concentrations of species are in units of mol•L^{-1}, although this formulation is problematic for insoluble substrates (*vide supra*). It is thus not uncommon to express the reaction components in units g•g^{-1}, mol•mol^{-1}, and mol•g^{-1}.

Unlike solution-phase enzyme rates that saturate with increasing $[S]$ but vary linearly with increasing $[E_T]$, cellulase reaction rates reach saturation with increasing $[E_T]$ and $[S_T]$. Thus, $[CE]$ must also be considered with respect to the mass balance for accessible binding sites on the substrate via:

$$\sigma[S_T] = [C] + [CE]$$

Equation 18

where σ is the binding capacity of the substrate, or the density of cellulase-accessible binding sites on the surface of cellulose in units of mol•g^{-1} and $[S_T]$ is described in g•L^{-1}. Equation 18 and Equation 15 are combined to produce the Langmuir isotherm in which the $[CE]$ can be modeled experimentally by titrating the reaction mixture with free enzyme:

$$[CE] = \frac{\sigma[S_T][E]}{K_d + [E]}$$

Equation 19.

The similarity between Equation 19 and Equation 4 under conditions of saturating enzyme ($E \to \infty$; $[CE] \to \sigma[S_T]$) suggests a qualitative relationship between the substrate-specific binding capacity σ and the two-dimensional conversion factor N_S^0 of the Gelb model. The Langmuir isotherm is the most commonly used equation to quantify adsorbed cellulases on a variety of cellulosic substrates.[88, 96, 110-112] We note further that K_d is a thermodynamic equilibrium constant, and kinetic limitations associated with establishing this equilibrium due to known tight-binding kinetics of cellulases may render a kinetic description of cellulase binding (Equation 16) more useful in the determination of $[CE]$;[113] it is typically assumed, however, that equilibrium is established within 90 min of enzyme introduction.[88]

Equation 19 can be combined with Equation 17 to generate a quadratic expression of $[CE]$:

$$[CE]^2 - [CE]\left(\sigma[S_T] + [E_T] + K_d\right) + \sigma[S_T][E_T] = 0$$

Equation 20

where the physically meaningful root satisfying the condition $0 \le [CE] \le \sigma[E_T]$ and $0 \le [CE] \le \sigma[S_T]$ is characteristic of the following limits:

$$[CE] = \begin{cases} \sigma[S_T],\ [E_T] >> \sigma[S_T] \\ [E_T],\ \sigma[S_T] >> [E_T] \\ \frac{[E_T]}{2} + K_d + \frac{\sqrt{[E_T]^2 + 4(K_d)^2}}{2},\ \sigma[S_T] \approx K_d \end{cases}$$

Equation 21.

When the boundary condition of $\sigma[S_T] \approx K_d$ is applied to simulate the native condition of $[C] \approx [CE]$, $[CE]$ remains a function of $[E_T]$, consistent with previously reported absorption characteristics of cellulases exhibiting dual saturation with substrate and enzyme.[88, 114-116] The second boundary limit from Equation 21, in which all enzyme is saturated with substrate sites, is the basis for the adaptation of the Gelb model to the Michaelis-Menten equation (Equation 10). However, the Michaelis-Menten equation is incompatible with dual saturation at high enzyme loading. Lynd et al. caution against the use of the substrate saturation limit for cellulase kinetics, since this limit is hardly applicable to native cellulase systems in which $[C]$ is typically not much higher than $[CE]$.[88] On the other hand, a more recent study of *Trichoderma veridi* cellobiohdrolase II Cel7A by Igarashi et al. suggests that the enzyme saturation boundary condition might lead to adverse over-crowding of the cellulose surface and lower specific activities of the enzyme.[117]

The equations above describe the kinetic behavior of cellulases and require titration cellulose surface sites with free enzyme in order to quantify the interaction

binding affinity (K_d) and substrate capacity ($\sigma[S_T]$). In practice, cellulase activity is far more complicated, and simple monitoring of free enzyme concentrations in tandem with cellulose hydrolysis rates to determine enzyme rate constants (k) yields values that vary by substrate or even laboratory. The irreproducibility of cellulase kinetics experiments is rooted in several factors: (i) cellulose is a heterogeneous substrate composed of multiple binding sites with varied crystallinity and cellulase-binding affinity; (ii) experimental conditions that may change from one laboratory to another appear to have a drastic impact on the determination of cellulase specific activity, as is apparent by comparison of the same enzyme by different groups;[88, 118] (iii) due to the long reaction times for cellulase reactions (hours to days), enzyme degradation often leads to time-dependent inactivation (aging) of the enzymes and time-dependent hydrolysis rates; and (iv) the intermediate and end products of endoglucanase-catalyzed cellulolysis act as inhibitors of cellulolysis, requiring evaluation of intermediate reactions to account for the variation in reaction kinetics:

$$r_{G_j} = f_{C\rightarrow G_j} r_C + \sum_i r_{G_{j,i}}$$

Equation 22.

Equation 22 describes the cumulative rate profiles for the formation of reaction intermediates such as oligosaccharides (G_j with DP = j) with positive rates ($+r_{G_j}$) when the oligomer is formed and negative rates ($-r_{G_j}$) when the oligomer is hydrolyzed by

cellulase. $f_{C \to G_j}$ denotes the fraction of cellulose hydrolyzed to oligosaccharides and r_C is the initial reaction rate as described in Equation 13. $r_{G_{j,i}}$ is the rate of formation of G_j by reaction i.[88] Of course, the application of Equation 22 further complicates the kinetics through the requirement of reaction monitoring tools to track the accumulation and depletion of soluble sugar intermediates over the time-course of cellulose hydrolysis.

1.4 Quantitative cellulosomics

Current understanding of the molecular genetics of cellulosome expression and evolution is extensive.[47, 119-122] Such studies qualitatively describe cellulosome activity on crystalline cellulose and suggest that fully-complexed cellulosomes degrade crystalline cellulose with synergistic improvement over non-complexed analogues. The total soluble sugar produced during cellulose hydrolysis was greater for scaffoldin-bound cellulases than for the individual free cellulases. This qualitative observation was quantified using "stimulation factors"—the ratio of soluble sugar produced by scaffoldin-bound cellulases divided by soluble sugar produced by the enzymes in the absence of scaffoldin—ranging in magnitude from 0.9 - 2.[121-123] To accurately describe the typical enhancement of cellulolytic activity, a more quantitative biophysical description of cellulosomal cellulolysis would be attractive in which both observable cooperativity of binding in the cellulase-cellulose interaction and the apparent synergy of hydrolysis could be related to cellulosome geometry.[124] In this section, we cover the current state-of-the art in cellulase structure and function analyses for cellulosomal systems.

1.4.1 Cellulosomal synergy: the targeting vs. proximity debate

To be or not to be… [complexed]? That is the question, for cellulases and scientists who study them. An early observation that inspired decades of research on the cellulosome was the realization that cellulosomal cellulases exhibited higher specific activity on crystalline and recalcitrant lignocellulosic substrates than did the corresponding cellulase mixtures from aerobic fungi. Cellulase researchers were already aware that cellulases in multi-component cellulase systems, such as the endo- and exoglucanase cocktails native to *Trichoderma reesie*, exhibited higher specific activities in conjunction with partner cellulases—1.7 fold increases in *T. reesei*—than did the individual cellulases working in isolation.[125, 126] It is generally accepted that the increase in activity was due to coopertivity between cellulases exhibiting orthogonal specificities. Similar studies with *C. thermocellum* cellulases found this enzyme system to be 159-fold more active in comparison to the free enzymes functioning in isolation.[127, 128] This level of synergy was not anticipated solely from the cooperativity between enzymes in a cellulase cocktail, and researchers began elucidating the functional role of the scaffoldin for cellulase activity enhancement in the cellulosome.[89]

At the turn of the century, J. P. Bélaïch, Bayer and their collaborators began a series of studies designed to explore the molecular basis of synergy in the cellulosome. Their initial hypotheses suggested two potential origins for cellulosomal synergy: (i) enzyme/ substrate proximity and (ii) substrate targeting.[121, 122] The proximity hypothesis

argued that a favorable spatiotemporal arrangement of endoglucanases near complimentary cellobiohydrolases yielded more efficient hydrolysis of cellulose, a substrate that changes during the course of hydrolysis as new sites for different enzymes are revealed during the course of the reaction. This hypothesis attracted much initial support based on the observations that *C. thermocellum* required a much lower enzyme load to completely hydrolyze crystalline cellulose than did the non-complexed cellulase system of *T. reesei*.[6] Enzyme targeting argued that the effective specific activity of cellulosome-integrated enzymes was greater than non-complexed cellulase mixtures due to an increased local concentration of cellulases at recalcitrant, crystalline sites on the cellulose matrix.

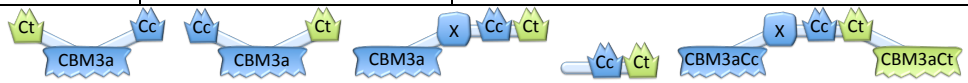
Using designer minicellulosomes (see section 1.1.3.2 Designer miniature cellulosomes), Bélaïch, Bayer, et al. reported that both effects seemed to play a role in cellulsomal synergy.[121, 122] They noticed that a pair of recombinant orthogonal cellulases (derived from the endoglucanase CelA and cellobiohydrolase II CelF of *C. cellulolyticum*) was 1.5-fold more active on crystalline cellulose when site-specifically complexed to a miniature scaffoldin bearing two orthogonal cohesin domains than when not complexed. When a scaffoldin CBM3a was added into the mincellulosome, the activity increased to 2-3-fold that of the non-complexed enzymes. It is interesting to note that no enhancement was observed for the complexation of two CelF molecules, although a 2-3-fold enhancement in activity was observed for two complexed CelA enzymes,

suggesting that endoglucanase activity is augmented to a greater extent than is exoglucanase activity.[121] Though previous work had already suggested that substrate targeting and/ or proximity might explain the scaffoldin-related synergy in the cellulosome,[129, 130] Fierobe's study was the first to separate targeting and proximity effects via the specific, modular incorporation of CBM and enzyme units. These results indicate that proximity and substrate targeting both likely contribute to cellusomal synergy, provided both endoglucanases and cellobiohydrolases are present. Contrary to previous studies that argued in favor of targeting, proximity seemed to be the more important factor.

In a follow-up study to the 2001 paper, Bélaïch et al used single-endpoint reducing sugar assays to evaluate the ability of 75 chimeric minicellulosomes to release soluble sugars (glucose and small cello-oligosaccharides) from avicel. The minicellulosomes comprised one of five designer scaffoldins and two of five *C. cellulolyticum* enzymes (CelA-GH5, CelC-GH8, CelE-GH9, CelF-GH48, or CelG-GH9) containing either a *C. cellulolyticum* or *C. thermocellum* dockerin domain to ensure site-specific inclusion on the chimeric scaffoldins. The scaffoldin-derived enhancements in enzyme synergy were quantified using *synergy factors* that defined and shown in Table 4.

It is important to initially consider the activity of the free-acting enzyme pairs. In general, family GH9 enzymes (CelG and CelE) are most active. When paired with itself, CelE yielded 86 μM soluble sugar from avicel after 24 hours at 37 °C. This was slightly

Table 4: Designer Cellulosome Synergy Factors (avicelase activity)[122]

Enzymes*	*Paired enzyme* activity (μM)	Free Enzymes SF#	*Scaf*1¶	*Scaf*2	*Scaf*3	*Scaf*4	*Scaf*5
5At/5Ac	28	0.9	2.5	3	2.3	1.3	1.8
5At/8Cc	28	1	3.3	3.6	2.8	1.5	1.6
5At/9Ec	76	1.2	2.9	3.1	1.9	1.5	1.6
5At/48Fc	49	1.2	3.6	4	3.1	1.9	2.4
5At/9Gc	39	0.9	7.2	7.3	5.8	3.5	4.7
9Et/5Ac	83	1.4	2.4	2.7	1.9	1.4	1.6
9Et/8Cc	69	1.3	2.9	3.3	2.2	1.3	2.2
9Et/9Ec	86	1	1.8	1.9	1.5	1.1	1.4
9Et/48Fc	77	1.2	2.2	2.2	2	1.8	1.5
9Et/9Gc	99	1.5	4.2	4.5	3.4	2.2	2.5
48Ft/5Ac	53	1.2	3	3	2.4	1.6	1.9
48Ft/8Cc	48	1.2	4.1	4.1	2.9	1.2	2.8
48Ft/9Ec	91	1.3	2.2	2.2	1.7	1.6	1.6
48Ft/48Fc	60	1.1	1.4	1.4	1.2	1.1	1.3
48Ft/9Gc	109	2	3.8	4	3.4	2.1	2.5
			Scaf1	Scaf2	Scaf3	Scaf4	Scaf5

Enzymes presented in format such that "5At" = GH5_CelA_*C.thermocellum. #: Synergy factor calculated as [Σ(enzymes together)/Σ(enzymes separate)]. ¶: Data represents synergy factors calculated as [Σ(enzymes on scaffoldin)/Σ(free enzymes)].

improved by pairing the enzyme with CelG or the CBH II CelF. The best combination of enzymes was GelG and CelF, exhibiting production of 109 μM soluble sugar, correlating to a 2-fold increase over the sum of the discrete enzymes. This exo/endo synergy appeared to be specific to the choice of enzymes and other CelF/EG combinations showed no significant enhancement over a pair of orthogonal CelF enzymes.

When the enzymes were complexed onto the designer scaffoldins (scaf1 – scaf5, Table 4), there was an enhancement in the enzyme synergy factor by 2-5 fold when a CBM3a was incorporated into the scaffoldin. Interestingly, the incorporation of enzyme pairs onto scaf4 led only to modest improvements in cellulose hydrolysis, with the notable exception of the CelA/CelG combination, which yielded a 3.9-fold enhancement in the synergy factor. This finding contrasts with earlier reports that enzyme proximity was the principle driving force for cellulosomal efficiency. Table 4 clearly demonstrates that CBM-containing scaffoldins exhibited the largest enhancements in enzyme synergy factors.

There appears, however, to be at least some bias for the choice of enzymes in addition to the geometry of the scaffoldin with respect to increased synergy factors and the origin of enhancement. The most striking enhancement is observed for the pairing of endoglucanases CelA and CelG. Analysis of this enzyme pairing on scaf4, which lacks the CBM, suggests an approximately equal role of enzyme proximity and substrate targeting for the increased synergy factor. This is evident as scaf4, which only allows for increased proximity, provides a 4X increase in the enzyme pair's synergy factor; while, the added ability to target cellulose through the CBM on scaf1 or scaf2 leads to an 8-fold enhancement over the non-complexed enzymes. A similar finding was apparent in the combination of endoglucanases CelE and CelG as a 1.3-fold increase in the synergy

factor was observed on scaf4. The synergy factors increased 2.8-fold when a CBM was added to the scaffoldin (scaf 1 and scaf2).

Enzyme targeting appeared to have a pivotal role in the pairing of exoglucanse with endoglucanase. CelF and CelG were essentially as active when complexed to scaf4 as when non-complexed, although incorporation of a CBM to target cellulose increased the synergy factor 1.9-fold. As this pair of enzymes is effective in the non-complexed form (synergy factor of 2 compared to the individually summed enzyme activities), it is unclear if the lack of proximity-related enhancement is actually due to substrate limitations. The dual-CBM scaf5 and alternate geometry of scaf3 generally reduced the enzyme synergy factors in comparison to scaf1 and scaf2, suggesting potential geometric constraints in these scaffoldins that limited enzyme activity.

Another study in 2005 stressed the importance of incorporating GH family 48 CelF and GH family 9 CelG into miniature cellulosomes in order to achieve optimal hydrolysis activity. Fierobe et al. paired these enzymes alongside a third orthogonal *C. cellulolyticum* endoglucanse in a trifunctional scaffoldin and determined complex activity against both cellulose and straw. The findings further illustrated the importance of the cellulosome archetype for proximity-related increases in enzyme synergy factors.[131] Other studies highlight the importance of enzyme targeting as equally important to enzyme proximity. Mingardon et al. observed the specific activity of CBM-ligated *C. cellulolycium* GH48 CelF and GH9 CelG on various cellulosic substrates. Their findings

clearly indicate that CBM-related targeting of the individual enzymes to the substrate increased the amount of soluble sugar released.[123] Bayer and co-workers reported similar findings using *C. thermocellum* GH48 and GH9-containing enzymes, highlighting the importance of cellulose-targeting rather than proximity for these families of cellulases.[132]

The studies of Bélaïch and co-workers are at best qualitative descriptors of the combined and, at times, conflicting contributions of enzyme proximity and substrate targeting as a function of scaffoldin geometry. Synergy factors from Table 4 are determined by endpoint assays of reducing sugar released from crystalline cellulose. A discussion of cellulase rate variations with respect to scaffoldin geometry is not presented, and the reducing sugar assays are highly variable indicators of indirect cellulase activity. Errors are large as Fierobe reported the need to repeat each cellulase experiment up to six times in order to reduce the data variance to 10%.[122]

1.4.2 Structural cellulosomics: the dissect and build strategy

Alongside the debate surrounding the origin of scaffoldin-related enzyme enhancement, cohesion-dockerin plasticity and specificity, and the genetic mining of cellulosomal organisms, a vast interest in the three dimensional structural determination of cellulosomal components has grown from the mid-1990's to the present day. Intial 3D characterizations began with the simplest components of cellulosomes: the type I cohesion[54, 55, 133] and dockerin domains of *C. thermocellum* and other Clostridia.[134]

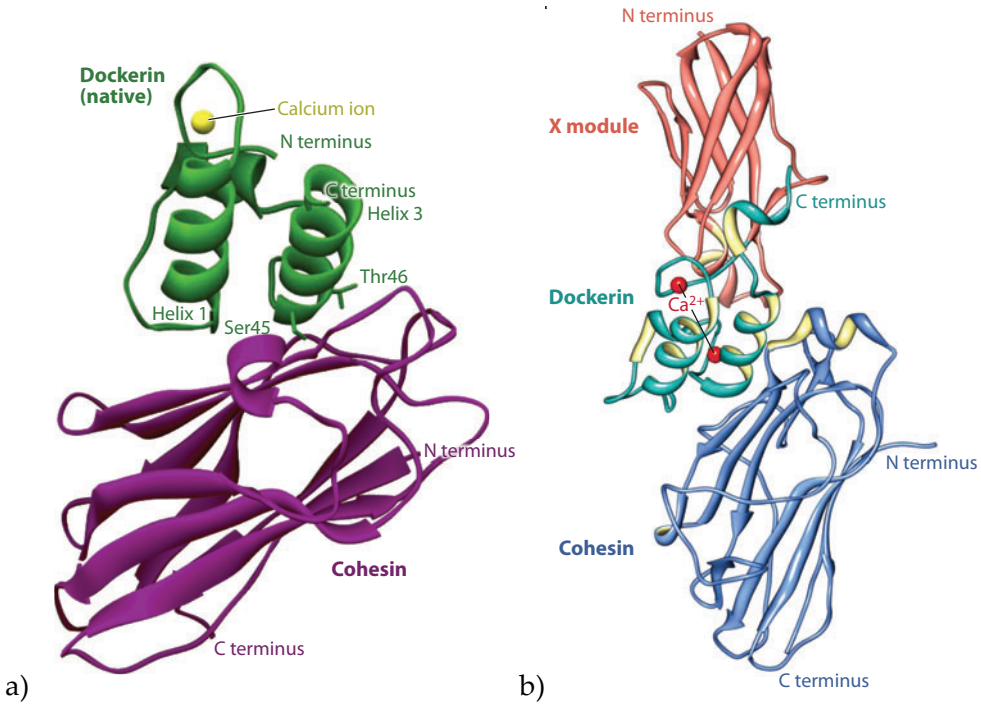


Figure 11: Clostridium thermocellum type I (a) and type II (b) cohesion-dockerin X-ray crystal structures (adapted from Fontes and Gilbert, 2010)[135]

The field continued to dissect the cellulosome into its constituents, providing structures of the CipA X-module of *C. thermocellum*, type II dockerin domains, CBM3a, type II cohesion and catalytic domains from multiple organisms.[68, 136-141]

The cellulosome components did not seem to tell the full story of cellulosome geometry as indicated by early electron microscopy,[12] and the field quickly shifted to crystallographic characterization of cellulosome complexes, beginning with the *C. thermocellum* type I cohesion-dockerin complex (Figure 11a).[49] An important study from Carvalho et al. provided valuable insight into the molecular detail of cohesion-dockerin specifities, particularly with respect to the role dockerin-based specificity determinants. As such, the structure laid a blueprint for cellulosome assembly and informed the design

of new cellulosome-based complexes. Adams et al. subsequently published a structure of the type II cohesion-dockerin complex that serves as the principle cell-membrane anchoring contact for the primary scaffoldin CipA (Figure 11b). Interestingly, this very tight interaction (K_a = 1.44 x 10^{10} M^{-1}) was found to involve both helices of the of the dockerin domain and also stabilizing contacts from the CipA X-module, suggesting a binding mode distinct from the type I dockerin-cohesin interaction, which requires only one of the dockerin helices for affinity.[52, 79, 142]

An impressive demonstration of the "dissect then build" strategy based on structural information came with the combination of cellulosome component crystal structures and low resolution small angle X-ray scattering (SAXS) data, computational modeling, and CryoEM images. In 2004, Hammel and co-workers used solution-phase, SAXS to monitor dynamic structural changes of the *C. cellulolyticum* GH48 CelF protein as it bound to a scaffoldin via a type I dockerin domain.[143] These results confirmed the existence of a two-domain structure with a flexible linker region between the GH domain and the dockerin. Interestingly, the linker compressed as the dockerin associated with the cohesion domain, suggesting an allosteric mechanism of scaffoldin-induced contraction, an effect that may play an active role in pre-organizing cellulases on substrates during cellulolysis.

Similar complexation-induced linker fluctuations during enzyme loading were also observed with high resolution CryoEM[13] and SAXS studies on multi-domain

scaffoldin proteins.[144] Unlike the contractions observed in the enzyme linker domains, the scaffoldin linker regions routinely display increased conformational heterogeneity upon enzyme binding. Likewise, X-ray crystal structures of *Acetivibrio cellulolyticus* cohesins with varying linker lengths presented linker domains with altered geometries, demonstrating linker flexibility in the context of other cellulosomes.[139] These observations further strengthen a structural mechanism for scaffoldin-induced pre-organization of enzymes on cellulose substrates for optimal cellulolysis in which the scaffoldin domain not only anchors the enzyme onto the cellulose through the CBM3a•cellulose association but also directs the enzymes to substrate sites for catalysis.[144]

In 2010, Smith and co-workers reported the crystallization of a trimodular scaffoldin (Doc2-X-Coh1) bound to a type-II cohesion (Figure 12).[141] The structure provides exquisite resolution of the scaffoldin geometry and depicts linker disorder within the scaffoldin similar to that previously described. The authors also observed homodimerization of the heterodimeric complex and confirmed dimerization by

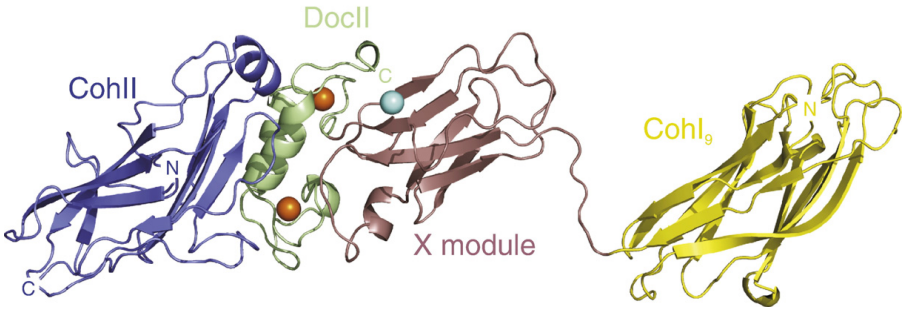


Figure 12: X-ray crystal structure of Clostridium thermocellum CipA and SdbA subunits (adapted from Adams et al, 2010)[141]

solution-phase sedimentation experiments. The dimerization was proposed as an alternate route towards cellulosome oligomerization (*vide supra*). Currie et al. followed this study with a combined X-ray crystallographic/ SAXS study of the same binary complex while bound to a type I dockerin domain derived from *C. thermocellum* GH9 CelD, forming a new ternary complex (Figure 13).[145] The complex forms an elongated structure similar to the binary complex and exhibits little evidence of scaffoldin linker contraction.[13] As before, the authors found evidence for homodimerization of the ternary

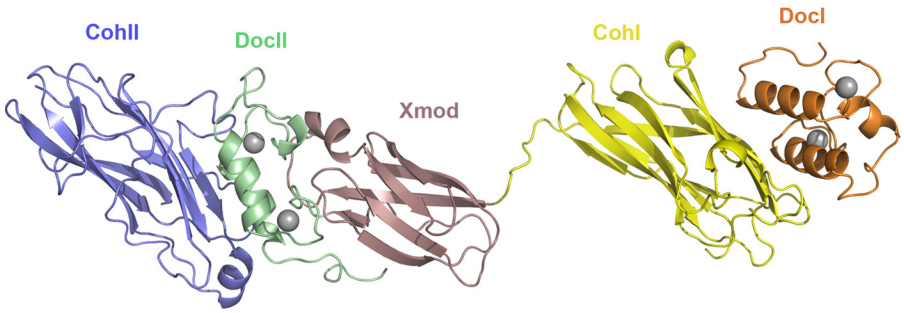


Figure 13: X-ray crystal structure of a ternary cohesion-dockerin complex from *C. thermocellum* (adapted from Currie et al, 2012)[145]

complex during crystallization, but solution-phase SAXS characterization of the complex suggested a monomeric, extended structure in solution.[145] In addition, four distinct conformations were visible in the crystal lattice, differing in the linker region between the type I cohesion and X module; this observation confirms previous suspicions that the scaffoldin linkers are highly dynamic. [139, 141, 143] In a more recent report, Currie and co-workers explored scaffoldin dynamics on the N-terminus of *C. thermocellum* CipA, opposite of the C-terminus X-Doc2 domain used to anchor the protein to SdbA.[146] Solution structures of GH8 CelA-bound miniature scaffoldins comprising two cohesions and a CBM3a domain described a highly dynamic complex with multiple conformations. The large degree of conformational heterogeneity is likely due to multiple flexible linker regions along the scaffoldin, though no apparent contraction within the CelA linker regions was observed upon cohesin-association.[143, 144]

Together these studies demonstrate the rapidly growing interest in the use of high-resolution X-ray crystallography in addition to SAXS to provide an atomistic illustration of cellulosome architecture. A long-term goal of the work in subsequent chapters is to continue the search for an understanding of cellulosomal synergy by means of building complexity from its parts.

2. CBM3a-carbohydrate association in solution and at interfaces explored with calorimetry and force spectroscopy

Cellulosomes contain a complex arsenal of carbohydrate recognition proteins that facilitate activity agains myriad carbohydrate substrates, ranging in complexity from simple oligosaccharides to highly recalcitrant, insoluble lignocellulose.[5, 111, 147-151] A subject of much debate centers on the role of these carbohydrate-binding proteins during catalysis by the full cellulosome. One argument, supported by the work of Lamed and Bayer points to so-called "targeting effects," whereby carbohydrate binding modules (CBMs) with different substrate specificities target cellulases to distinct cognate substrate sites with the associated catalytic activity.[152] As described in chapter 1, substrate targeting plays an integral role in the overall enhancement of cellulosomal cellulolysis. The cellulosome contains an enormous array of carbohydrate-binding proteins that directs the cellulosome to a diverse set of substrates and provides a potential evolutionary advantage by enabling hydrolysis of recalcitrant cell wall polysaccharides.

2.1 The unique binding modes of cellulosomal scaffoldin CBM3a

CBMs were originially classified as A, B, or C. This classification scheme groups CBMs according to substrate specificy of cellulose binding domains (sometimes referred to as CBD). Type A CBMs bind crystalline polysaccharides, type B CBMs bind amorphous polysaccharide chains, and type C CBMs bind small water-soluble sugars. A

more recent classification scheme based on sequence similarity groups CBMs into 39 numeric families, though often the original domain classification is also included to inform the known binding capacity of a CBM (i.e. CBM3a is a family 3 CBM that binds crystalline cellulose).[153]

CBM diversity within the cellulosome is immense due to the incorporation of type B and C CBMs within dockerin-containing cellulosomal endoglucanases in addition to scaffoldin-borne CBM3a. The cellulosomal enzyme-borne CBMs are functionally similar to the analogous CBMs found in non-cellulosomal endoglucanases. As such their roles are thought to mirror those of the non-cellulosomal CBM, namely binding to amorphous regions of cell wall polysaccharides to direct cellulase-catalyzed cellulolysis.

Type 3a CBMs are uniquely cellulosomal. These domains exist only within the non-enzymatic scaffoldin component of the cellulosome and are thought to bind only to insoluble beta-glucans, such as cellulose, chitin, and chitosan. CBM3a is considered essential for the efficient saccharafication of crystalline cellulose, but its role in the degradation of amorphous and soluble cellulosic substrates has been largely ignored.

To date, three Clostridial scaffoldin CBM3a domains have been crystallized;[154, 155] the structures provide a plausible rationale for the preference of CMB3a domains for crystalline polysaccharides. The binding site of these CBMs is formed from a large, flat, β-sheet in the protein. The binding site presents several canonical sugar-binding residues at the binding interface. Unfortunately, no substrate-associated structures have

been reported for this family of CBMs; such a structure is unlikely to appear in the near future due to the recalcitrance of cellulose.

Torno and co-workers have speculated on the roles of specific aromatic and charged residues for sugar binding along the lower CBM3a β-sheet.[154, 155] The investigators hypothesized a unique binding mode in which the binding residues do not complex multiple pyranose rings of a single chain but, rather, associate with pyranose moieties of adjacent cellulose chains. This binding mode has interesting implications for cellulosome attachment to cellulose. Two corollaries flow from the proposed binding mode: 1) CBM3a•cellulose affinity is a function of cellulose surface crystallinity, and decreases with increasing amorphous character of the substrate; and 2) CBM3a•cellulose binding is readily disrupted by single chains of cellulose or soluble cello-oligosaccharides that occupy a sub-site along the binding interface. These hypotheses are largely untested, although, qualitatively, CBM3a does not appear to exhibit significant affinity for amorphous cellulose.

We aim to develop a molecular basis for CBM3a affinity due to several motivating factors. A biophysical characterization of CBM3a•ligand interactions is necessary to illuminate the molecular basis of substrate targeting in the cellulosome. In addition, cellulase kinetic models (see chapter 1) highlight the need to quantify CBM binding to cellulose substrate and product surface sites as well as soluble sugar products of cellulolysis. All literature reported to date regarding the interaction of cellulosomal

proteins with insoluble cellulosic substrates is limited to qualitative observations of binding.[122, 156] To our knowledge, no studies regarding the interaction of type 3a CBMs with soluble forms of cellulose and simple cellodextrins have been reported. This deficit is surprising, since such studies would provide insight into the intricate interplay of protein-carbohydrate interactions within the cellulosome during the complex process of cellulolysis. In addition, a comparison of binding free energies of CBM3a binding to different cellulosic ligands would provide valuable insight into the evolution of protein structure that imparts substrate specificities across CBM families.

In this chapter, we detail an experimental approach to the systematic evaluation of the ability of a *Clostridia cellulolyticum* CBM3a domain to bind water-soluble cellodextrins. We later explore the specific interactions of this domain at the protein-cellulose interface using force spectroscopy. Prerequisite to quantifying CBM3a•soluble sugar affinities is the synthesis of water-soluble cello-oligosaccharides and the production of recombinant CBM3a-containing proteins.

2.2 Hybrid approach to the production of cello-oligosaccharides

Following the synthetic procedure of Redlich et al., peracylated cello-oligosaccharides, or *cellodextrins*, were generated in mg to g quantities up to DP_{12} via the Lewis acid-catalyzed pivaloylysis of cellulose triacetate.[157] This approach allows for the efficient acquisition of small to medium-sized cellodextrins, species that are not

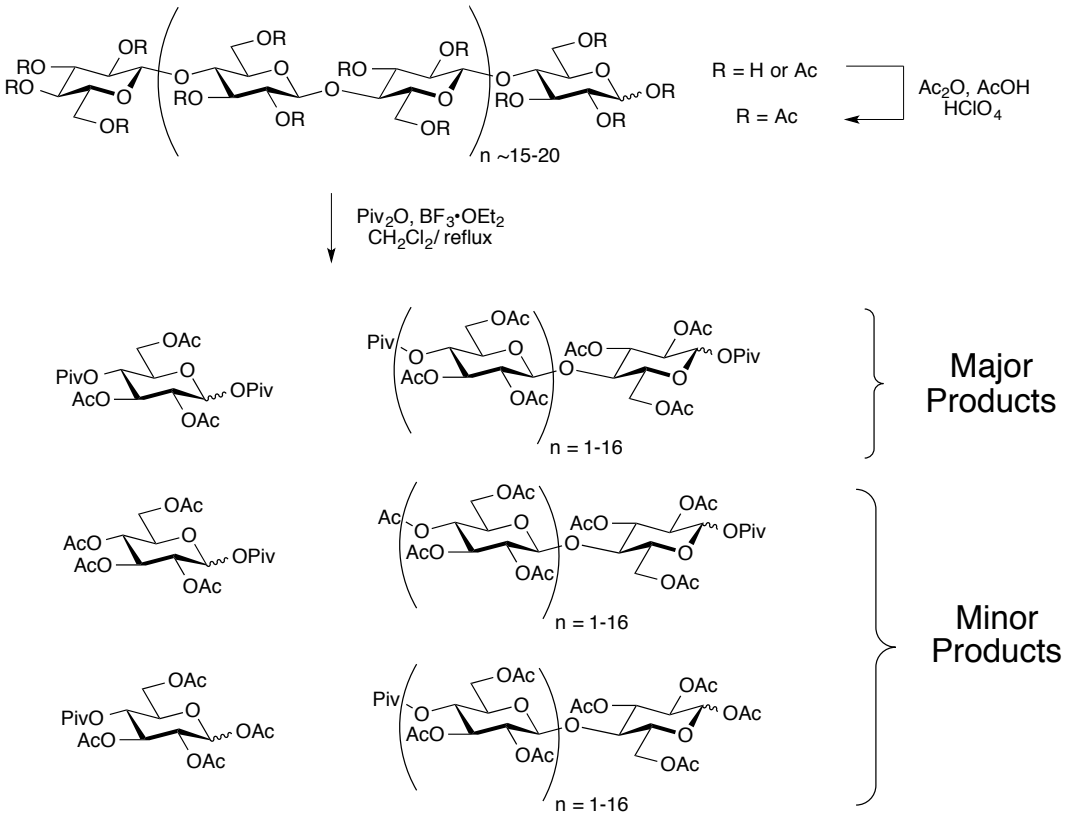
available commercially in the quantities and purities required for our biophysical studies.

Production of peracylated and underivatized cellodextrins is described in the following sections. Due to the varying water solubility of cellodextrins, we utilized a "mixed purification" protocol whereby insoluble cellodextrins are isolated in the acylated form in organic milieu, while small cellodextrins (DP_1 – DP_7) were obtained as a mixture, de-protected, and then purified in aqueous milieu. Purification of acylated and underivatized cellodextrins was achieved using a combination of normal phase flash chromatography, reversed phase high performance liquid chromatography (HPLC), calcium affinity chromatography, and size exclusion chromatography.[158, 159]

2.2.1 Pivaloylsis of cellulose triacetate

Modeled after a milder pivaloylysis alternative to the acetolysis of cellulose acetate, fully acylated cellodextrins were obtained via the process of Scheme 1. Cellulose triacetate was prepared from commercial cellulose acetate as previously described[157] and the product was depolymerized by treatment with boron trifluoride diethyl etherate and pivalic anhydride.[157] The pivaloylysis was monitored over the course of 25 hours by maxtrix-assisted laser desorption ionization mass spectrometry (MALDI-MS). The mild nature of the depolymerization is apparent in the MALDI-MS spectra (Figure 14). The accumulation of intermediate cellodextrin acetates ($3 \leq DP \leq 17$) over the course of 18-25

hours suggests that the optimal reaction time to acquire cellodextrins of $DP_{3\text{-}8}$ is near 20 hours.



Scheme 1: Production of cellodextrins by pivaloylysis of cellulose triacetate

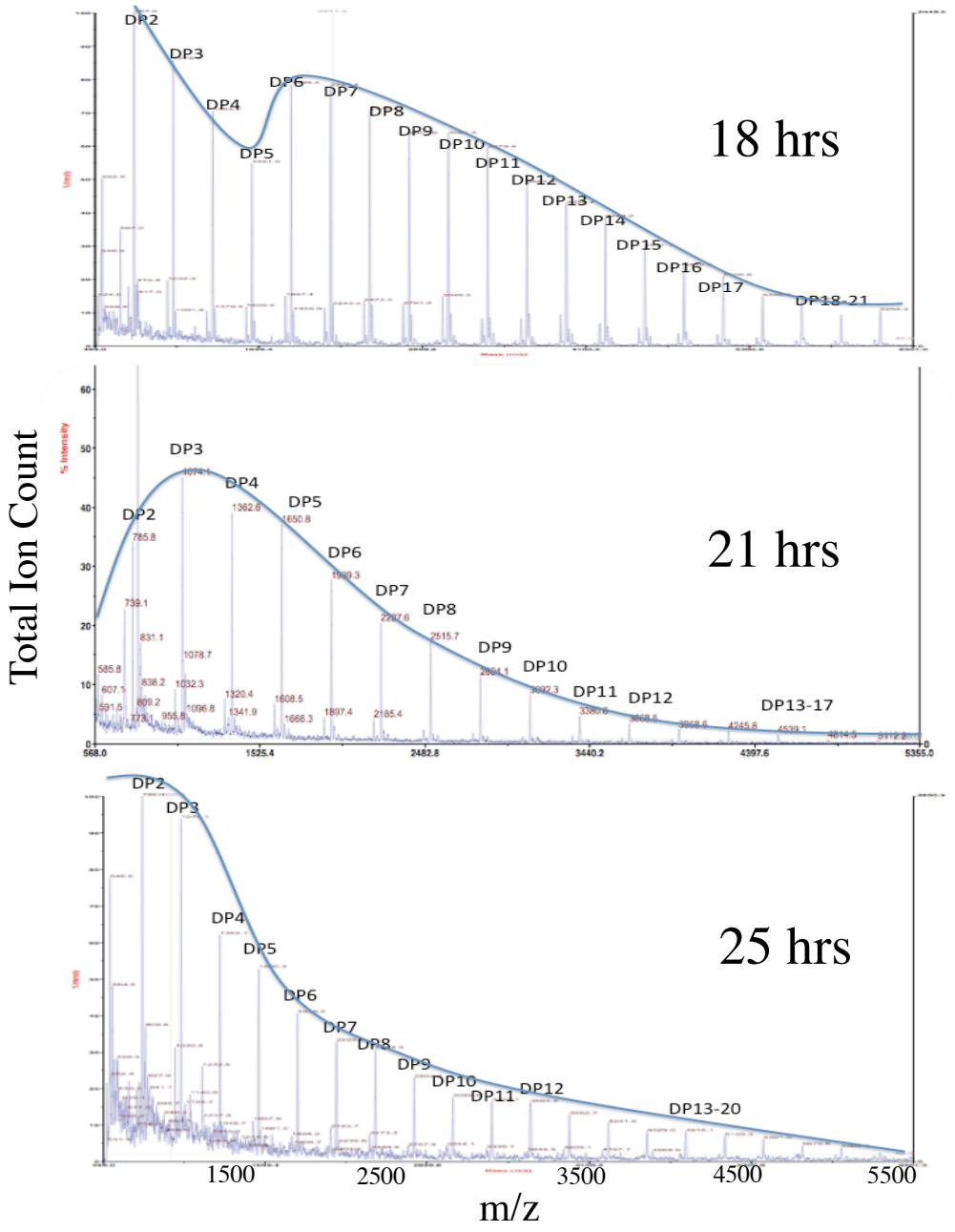
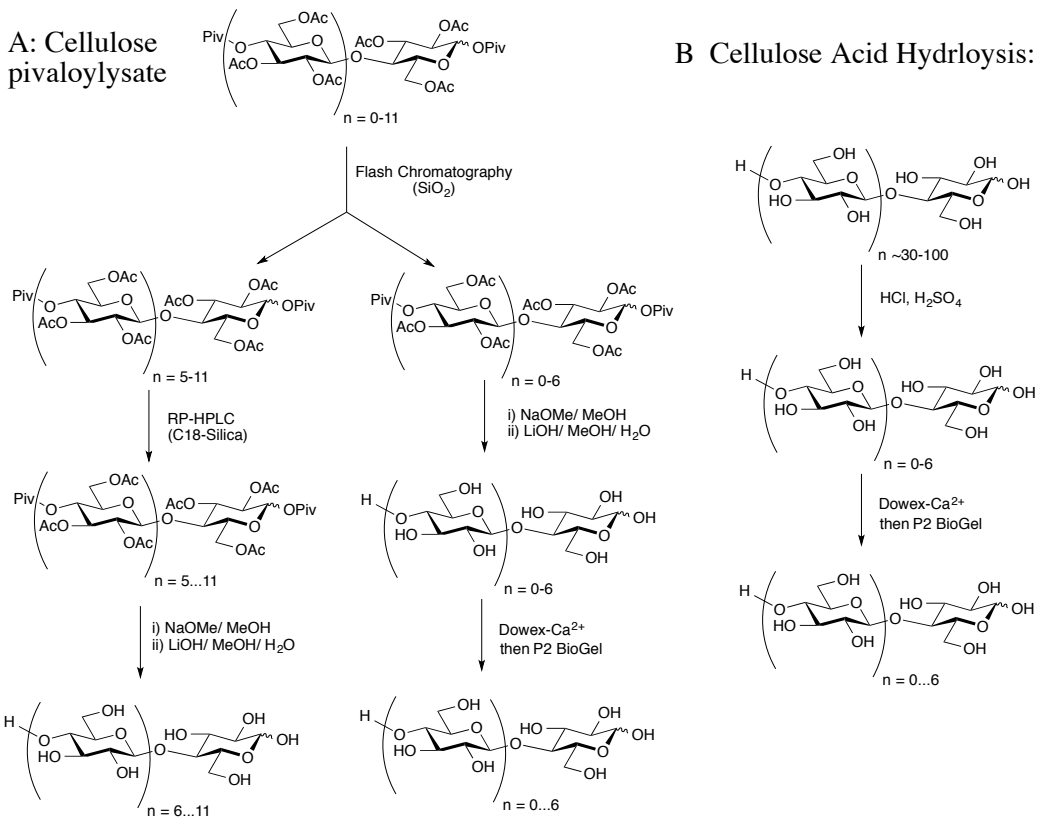


Figure 14: MALDI-TOF analysis of the pivaloylysis of cellulose triacetate

2.2.2 Purification of per-acylated and underivatized cellodextrin mixtures



Scheme 2: Mixed purification protocol for cellodextrins

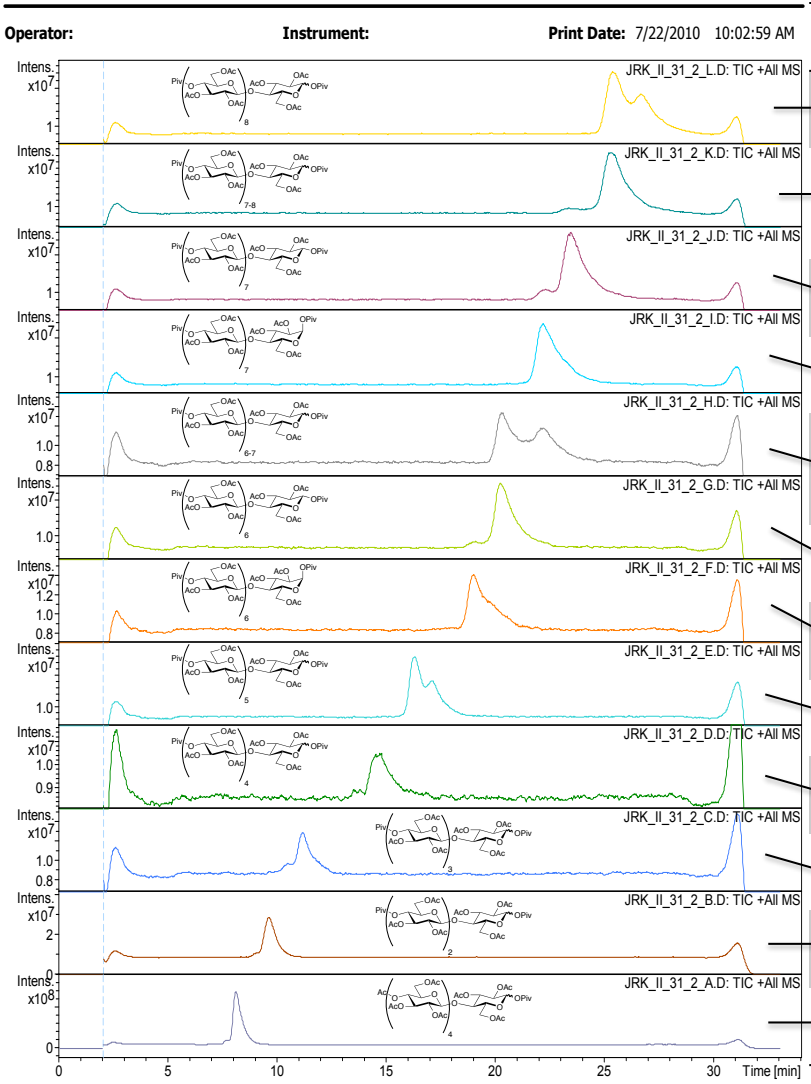
Cellulolse pivaloylysate primarily comprised a mixture of C1,C4'-pivaloyl-capped cellodextrins of DP = 1-12. The purification of this complex mixture is outlined in Scheme 2A. The crude pivaloylysate was first partitioned over silica gel to yield large (DP > 5) and small (DP <8) cellodextrin acylate mixtures. The small cellodextrin mixture was then subject to batch deprotection under Zemplen conditions followed by saponification of the pivaloyl groups to yield water-soluble cellodextrins ranging from glucose to celloheptaose. As cellodextrins larger than celloheptaose are no more than

sparingly soluble in water, cellodextrin acetates of DP > 7 were purified separately using reversed-phase chromatography and de-protected individually.

An alternative method for the rapid production of small cellodextrins (DP_{3-6}) was reported by Zhang et al. (Scheme 2B).[158] This method utilizes a mixed acid hydrolysis medium composed of concentrated sulfuric (>98%) and hydrochloric acids (37%) and produces predominantly cellobiose and glucose. However, at short reaction times (3-5 hours) substantial amounts of intermediate hydrolysis products, including cellodextrins ranging in DP from 3—6, can be isolated. This method was used in conjunction with the previously described pivaloylysis to rapidly produce cellodextrins up to DP_6. As before, the poor solubility of large cellodextrins ($DP_{>6}$) in water renders the mixed acid hydrolysis method inferior in comparison to the pivaloylysis route for production of large cellodextrins.

2.2.2.1 Reversed-phase high-performance liquid chromatography of peracylated cellodextrins

Protected cellodextrin mixtures from pivaloylsis were purified over octadecyl-modified silica via reversed-phase HPLC, producing single oligomers as anomeric mixtures, though some of the anomers were isolable as single diastereomers (Figure 15). The purification was achieved by application of a near-isocratic gradient of acid-modified acetonitrile in water (60-64% MeCN: Water with 0.02% formic acid) over 30 minutes (see chapter 5). Such a gradual gradient or isocratic elution over C18-silica is characteristic of the RP-HPLC purification of cellodextrin acetate mixtures.[160]



ion	m/z
$[G9+2Na]^{2+}$	1413
$[G8+2Na]^{2+}$	1268.5
$[G9+2Na]^{2+}$	1413
$[G8+2Na]^{2+}$	1268.5
$[G8+2Na]^{2+}$	1268.5
$[G7+2Na]^{2+}$	1125
$[G8+2Na]^{2+}$	1268.5
$[G7+2Na]^{2+}$	1125
$[G7+2Na]^{2+}$	1125
$[G6+Na]^{+}$	1937
$[G5+Na]^{+}$	1649
$[G4+Na]^{+}$	1361
$[G3+Na]^{+}$	1073
$[AG5^{*}+Na]^{+}$	1608

Figure 15: Reversed-phase LCMS of peracylated cellodextrins

The separation of piv-capped cellodextrin acetates was monitored by off-line liquid-chromatography coupled to electrospray ionization mass spectrometry (LCMS). Total ion chromatograms are depicted in Figure 15. The value of LCMS for the characterization of complex oligosaccharide mixtures cannot be overstated. In addition to confirming product formation and homogeneity, the chromatograms facilitate

identification of cellodextrin anomers and reaction byproducts, such as the C4'-acetyl pentasaccharide (AcG5) shown in Figure 15, which elutes before the piv-capped trisaccharide (G3). This latter finding, though consistent with the pivaloylsis mechanism, was somewhat unexpected, since Redlich et al. provided no indication that these reaction products were produced in significant amounts.[157] The omission may be a result of the use of one-dimensional ^{1}H- and ^{13}C-NMR rather than LCMS to characterize the reaction products. Indeed, C4'-acetate vs. C4'-pivalate reaction products would present nearly-identical NMR spectra, especially when recording the spectra of large cellodextrins and complicated diastereomeric mixtures.

2.2.2.2 Tandem calcium-affinity and size exclusion chromatographies for the separation of underivatized cellodextrins.

The mixed purification and hydrolysis strategy outlined in Scheme 2 required batch deprotection and subsequent purification of the underivatized cellodextrin product mixture. Zhang and Lynd reported the synergistic combination of calcium affinity and size exclusion chromatographies as appropriate for this task.[158] The binding capacity of cellodextrins to Ca^{+2} decreases with increasing degree of polymerization, thus larger cellodextrins eluted from a Ca^{+2}-loaded DOWEX resin prior to smaller oligomers. Passing the eluent directly over the high resolution size exclusion resin P2-Biogel, which efficiently fractionates biomolecules in the 100-2000 Da size regime, further refined the initial separation achieved using calcium affinity chromatography.

Thus, with only minor adaptations we were able to quickly and routinely obtain large quantities (0.1- 1 gram) of single cello-oligomers.

2.3 Heterologous over-expression of recombinant clostridial cellulose-binding domains in Escericia coli

High expression, or "over-expression," of recombinant DNA plasmids is now commonplace in biochemical research.[161] The process involves the synthesis of an *expression vector*, which is a circular DNA fragment usually derived from a commercial pET-based plasmid that contains the following minimal components: i) one or more resistance cassettes that confer anti-bacterial resistance; ii) the T7 bacteriophage RNA-polymerase promoter sequence; iii) a *lac* operator which blocks the transcription process prior to the addition of lactose or isopropylthio-β-D-galactopyranoside (IPTG), and v) a polylinker region containing unique restriction endonuclease sites for site-selective incorporation of foreign genes.

The bacterial host is typically modified (i.e. the *E. coli* strain BL21(DE3)) to contain the T7 RNA polymerase gene under the control of the repressed *lac* operon. T7 is a robust bacteriophage RNA polymerase not native to the bacterial host. The only promoter site for T7 polymerase is thus found on the pET-based expression vector.

The process of host transformation and protein expression relies on a series of internal "checks" that ensure optimal growth of only the desired *E. coli* transformants expressing the gene of interest. First, the pET vector confers anti-bacterial resistance (i.e. the production of beta-lactamase to degrade the beta-lactam anti-biotic ampicillin) to the

host. Upon transformation of the host with the foreign DNA, the bacteria is grown in the presence of ampicillin (or another antibiotic); under these conditions, only hosts containing the gene of interest survive.

Next, transcription of the recombinant gene on the pET vector is induced by addition of lactose or IPTG, which disrupts the interaction of repressor proteins of the *lac* operon. This repression allows the simultaneous expression of T7 RNA polymerase from the bacterial genome and subsequent binding and transcription of the pET vector. During this process, the native protein translation protein machinery of the host is saturated by the large amounts of foreign RNA produced, quickly leading to high yields of the target protein.

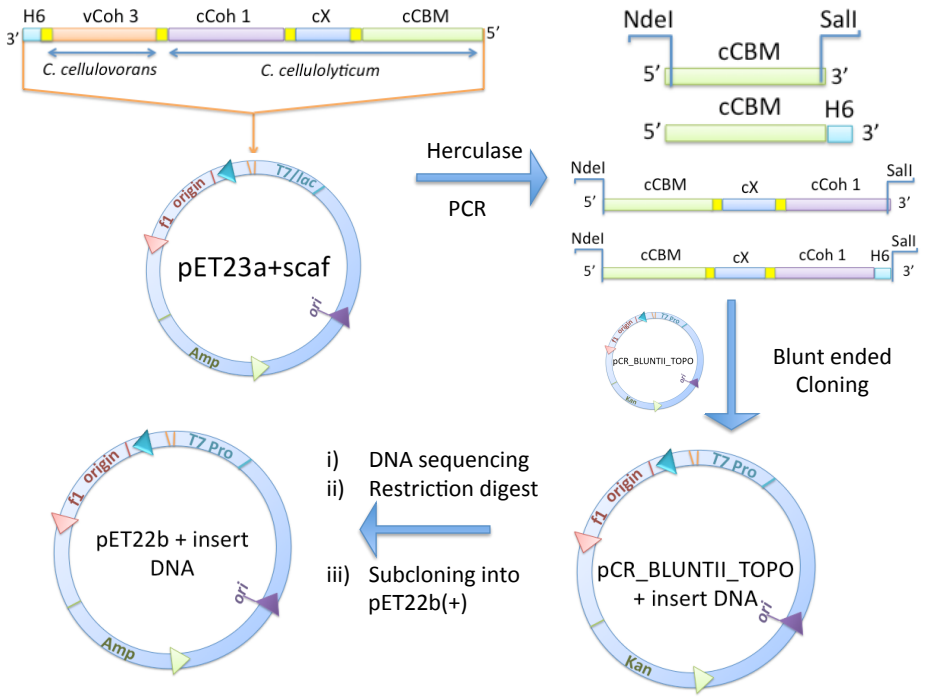
The heterologous overexpression of proteins in *E. coli* is a powerful tool but is often plagued by numerous problems, including protein stability, protein-host compatibility, and protein solubility. The protein and/or expression plasmid itself is often toxic, requiring rapid induction (over ~3 hours) in order to collect the *E. coli* before substantial apoptosis has occurred. Alternatively, high expression rates achieved via the use of large amounts of IPTG or high expression temperatures (30-37 °C) might lead to aggregation, precipitation or mis-folding of the protein, processes that often lead to formation of insoluble protein inclusion bodies that must be denatured and reconstituted into native form. Monitoring protein expression over various IPTG

concentrations, expression temperatures, and expression time frames is often required to minimize these problems.

Heterologous expression of Clostridial CBM3a protein domains and the recombinant protein miniCipC in *E. coli* was successfully performed at high yields (up to 0.1-0.2 g/L). We expressed the CBM3a domain of the *C. cellulolyticum* CipC primary scaffoldin in order to detail the binding mode of this CBM to water-soluble cellodextrins and insoluble cellulose. We hypothesized that the modular nature of the scaffoldin domain may play a role in substrate recognition, and we also expressed the miniCipC scaffoldin using the approach of Pages et al.[162]

2.3.1 Design and synthesis of recombinant CBM3a-encoding plasmids

pETCBM, pETCBMHis, pETminiCipC, and pETminiCipCHis are pETscaf-derived protein expression vectors constructed via enzymatic insertion of CBM3a or miniCpC-encoding genes into the pET22b+ expression plasmid with either the addition or omission of C-terminal stop codon, leading to genes without or with a hexahistidine affinity tag at the C-terminus, respectively. pETscaf is a pET23a-based vector containing a gene designed by our collaborators in the Fierke lab at the University of Michigan at Anne Arbor and synthesized by IDT DNA Technologies. The *scaf* vector was designed to encode a tetramodular miniature scaffoldin protein analogous to Scaf 3 reported by Fierobe et al.[121] The reverse primers for the miniCipC-encoding DNA carried 15 extra

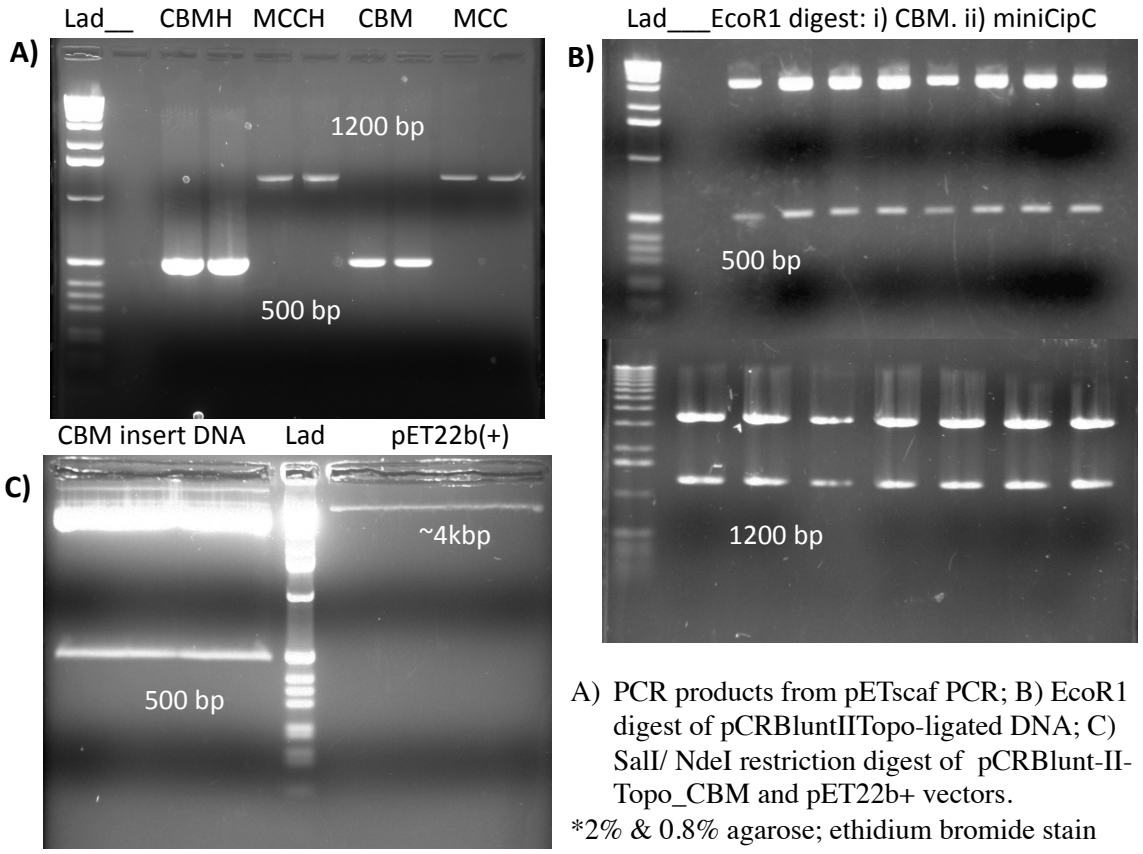


Scheme 3: Construction of pET22b-based plasmids

codons to allow incorporation of a LKVAV sequence at the C-terminus, ensuring proper globular fold of the mature protein.[134]

The genes were constructed under general cloning procedures as depicted in Scheme 3 and described in detail in Chapter 5. Gene synthesis was followed by agarose gel (Figure 16). Initial PCR amplification of the *cmb3a* or *minicipc* DNA regions of the synthetic vector pETscaf was performed using the high-fidelity, thermostable Herculase DNA polymerase from Agilent Technologies. The blunt-ended PCR products were subcloned into the commercial high-copy sequencing vector using the Blunt-II-Topo cloning technology from Invitrogen and transformed into the XL10(Gold) ultracompetent *E. coli* strain from Agilent Technologies. The pCR-BluntII-Topo-based

plasmids were amplified in *E. coli* and purified using Qiagen plasmid miniprep kits. The plasmid DNA was subjected to restriction digest by SalI and NdeI endonucleases to excise the *cbm3a* and *minicipc* genes. Ligation of the excerpts into a pre-digested and phosphatase-treated pET22b+ expression plasmid was performed using fresh T4-ligase, resulting the construction of pETCBM and similar plasmids.



A) PCR products from pETscaf PCR; B) EcoR1 digest of pCRBluntIITopo-ligated DNA; C) SalI/ NdeI restriction digest of pCRBlunt-II-Topo_CBM and pET22b+ vectors.
*2% & 0.8% agarose; ethidium bromide stain

Figure 16: Agarose Gel Electrophoresis of CBM and MiniCipC plasmid synthesis

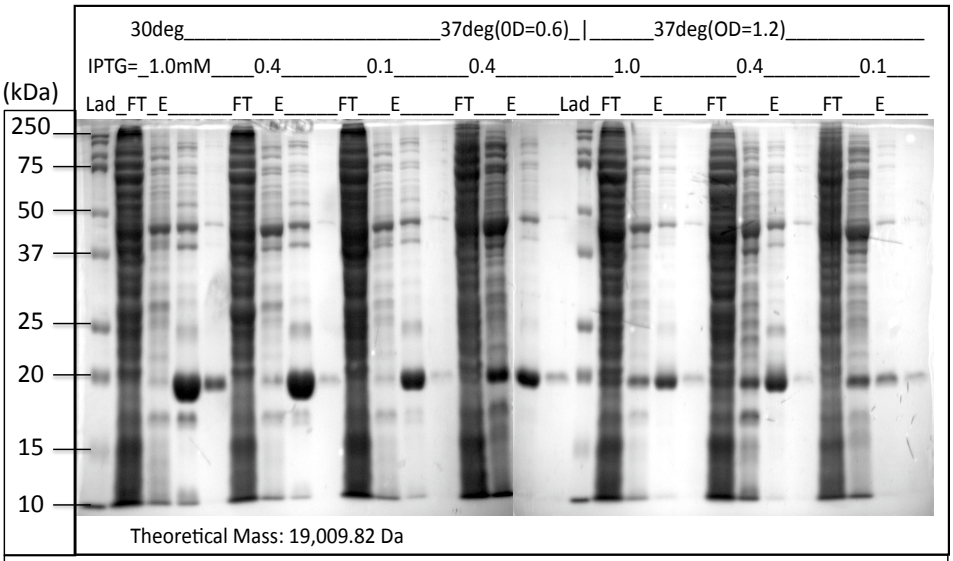
MiniCipC-encoding DNA within the pETscaf vector displayed a point mutation whereby an A→C mutation caused an N245T mutation. This mutation was reversed

using quick-change PCR as described in chapter 5 to yield the appropriate miniCipC expression vectors pETminiCipCTΔN and pETminiCipCTΔNHis.

2.3.2 Expression and purification of *C. cellulolyticum* CBM3a

Clostridium cellulolyticum CBM3a (from pETCBM or pETCBMHis) was expressed under control of the T7/lac-operon system in BL21(DE3) *E. coli* cells at varying temperatures, concentrations of IPTG, and expression times. A qualitative comparison of the amount of His-tagged protein expression under different conditions was performed via the analytical-scale lysis of protein-filled cell pellets and subsequent crude immobilized-metal affinity chromatography (IMAC) of the lysate under denaturing conditions. Sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the small-scale IMAC purifications is shown in Figure 17. The results indicate that CBM3a expression is most strongly influenced by temperature, expressing best at 30 °C. Likewise, higher [IPTG] produced improved yields.

A detailed description for CBM3a purification by fast protein liquid chromatographic (FPLC) is provided in the experimental chapter 5. Briefly, CBM3a was isolated from the water-soluble fraction of crude cell lysate via cellulose affinity chromatography and subsequently purified to homogeneity (as analyzed by SDS-PAGE) via size exclusion chromatography using an HR-S100 gel filtration column (GE Healthcare). Chromatographs and PAGE gels are shown in Figure 18.



*OD = Optical Density at 600 nm; Lad: Protein MW Markers; FT: column flow through; E: protein elution at low pH; Gel: 12% acrylamide and Coomassie stain.

Figure 17: Expression analysis of His_6-tagged CBM3a

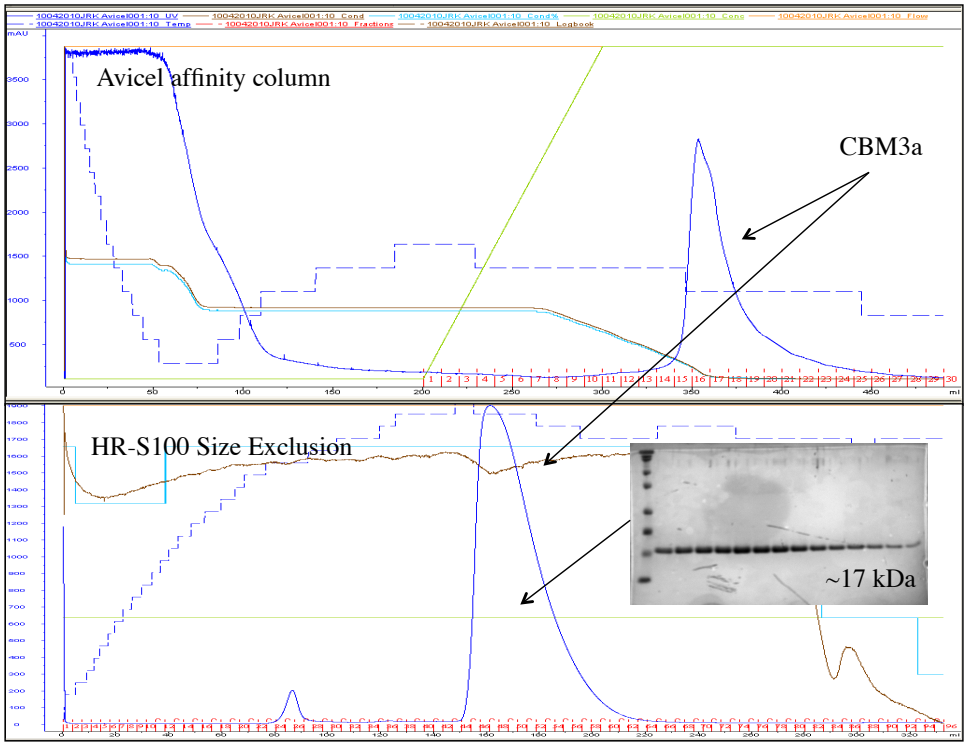


Figure 18: CBM3a purification by cellulose affinity and size exclusion chromatographies

2.3.3 Expression and purification of the miniature scaffoldin miniCipC

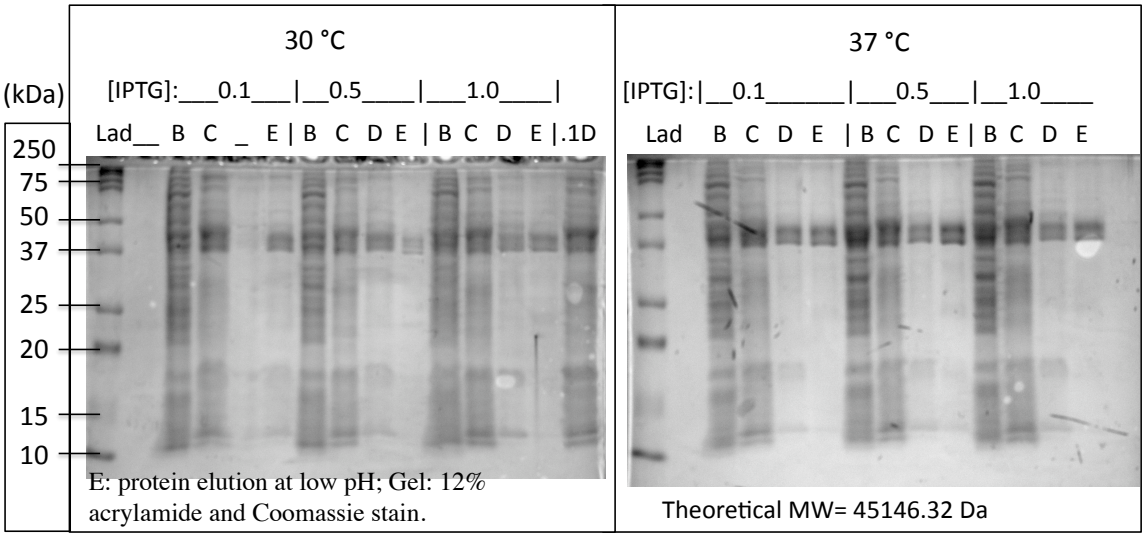


Figure 19: Expression conditions for miniCipC

The chimeric scaffoldin "miniCipC" was heterologously expressed from pETMCCTΔNHis expression vector in *E. coli* under the control of the T7 RNA polymerase/ lac operon system in BL21(DE) cells. Specific expression conditions are listed in the experimental chapter 5. A qualitative comparison of the amount of His-tagged protein expression under different conditions was performed via the analytical-scale lysis of protein-filled cell pellets and subsequent crude IMAC pull-down from the lysate under denaturing conditions. SDS-PAGE analysis of the small-scale IMAC purifications is shown in Figure 19. The results indicate that protein expression does not vary significantly with temperature or [IPTG]. Closer inspection of the expression time-course indicated that optimal miniCipC expression occurred at 0.4 mM IPTG and 30 °C during seven hours.

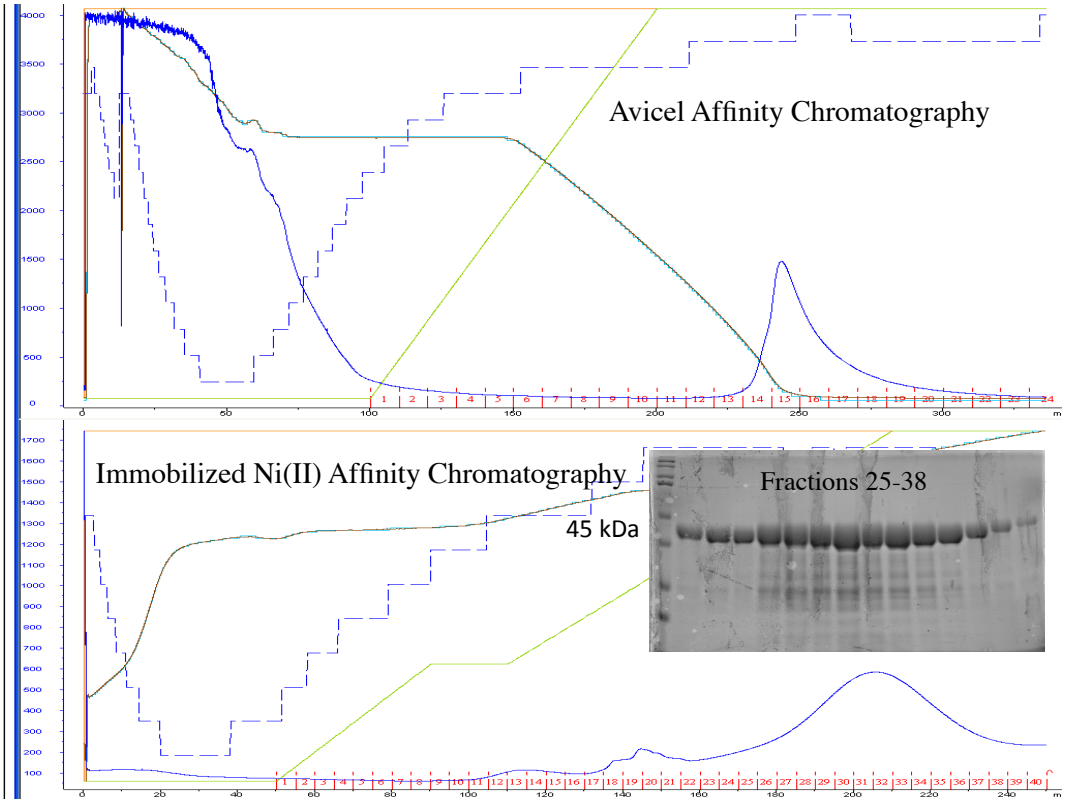


Figure 20: Affinity purification of miniCipC over cellulose and Ni^{2+}-NTA resin

MiniCipC was isolated from the water-soluble fraction of crude cell lysate via cellulose affinity chromatography and subsequently purified (>90% pure, as analyzed by SDS-PAGE) via a Ni(II)-loaded IMAC column with elution to 300 mM imidazole over 20 column volumes; details of the purification are provided in chapter 5. Chromatographs and PAGE gels are shown in Figure 20.

2.4 Does CBM3a bind soluble sugars?

As described above, type A CBMs bind crystalline polysaccharides such as cellulose, chitosan, or chitin. The CipC CBM3a domain from *C. cellulolyticum* is structurally designed to accommodate the large, flat surface of crystalline cellulose. The

binding site was hypothesized to accommodate adjacent cellulose chains, rather than the linear glucans that comprise the cellulose surface. To the best of our knowledge, however, no studies that explore the effects of soluble cellodextrins on the binding of any type CBM3a to crystalline cellulose have been reported. Gilbert and co-workers demonstrated that scaffoldin CBMs bind different forms of cellulose with a range of affinities.[163] They reported that a family 29 CBM found on a cellulosomal scaffoldin protein of the anaerobic fungi *Piromyces equi* efficiently binds polymeric cellulose (ΔG = -7.6 kcal•mol^{-1}) but also displays a reduced capacity for soluble cellodextrins (-4 < ΔG < -6 kcal•mol^{-1}).

We sought to determine whether CBM3a binds soluble sugars in an attempt to develop a molecular mechanism for CBM3a-cellulose targeting. Similar to CBM29 described above, we hypothesized that cellodextrins may display a comparable binding affinity for CBM3a as compared to insoluble cellulose, which would facilitate cellulosome release from cellulose as soluble cellodextrins are released in large concentrations at the site of cellulolysis and compete for the CBM3a binding sites. Presumably, cellulosome would shift to an area of lower cellodextrin concentration, re-bind the cellulose matrix, and continue cellulolysis. Initially, direct interactions of CBM3a with cellodextrins were explored using istothermal titration calorimetry.

2.4.1 CMB3a-cellodextrin binding probed with Isothermal Titration Calorimetry

CBM3a solutions were titrated with soluble cellodextrins of increasing DP from cellobiose to cellopentaose in a VP model ITC. The isothermal titration microcalorimeter was introduced in 1989 by Wiseman and co-workers[164] and has since become the gold standard for biophysical evaluation of intermolecular interactions, solvation thermodynamics, and enzyme kinetics.[165-167] Calorimetric evaluation of protein-carbohydrate affinities is a valuable tool and a long-standing interest of the Toone laboratory.

The calorimeter functions by measuring temperature differences between a sample cell containing a binding partner of the assayed reaction (usually the macromolecule) and a reference (usually containing buffer or solvent) in an adiabatic jacket. As injectant is added to the sample cell, heat is evolved or absorbed by a chemical reaction between or dilution of the analytes. The instrument continually applies more or less power (dQ/dt) to a sample cell in order to maintain an isothermal relationship between the sample cell and a reference cell in the course of the titration. At the end of the titration, individual heat signals are integrated over time yielding the heat or absorbed or produced at each step of the titration (ΔQ), in turn related to the enthalpy of binding in a binding titration via the expression

$$\Delta Q = n\Delta[MX]\Delta H^{\circ}V_0$$

Equation 23,

where n is the binding stoichiometry, ΔH° is the molar binding enthalpy in kcal•mol^{-1}, V_0 is the cell volume (1.43 mL in a typical VP-ITC), and [MX] denotes the complex of macromolecule M and ligand X described the binding equilibrium M + X $\rightleftharpoons$ MX. Given an equilibrium constant K_{eq} defined as

$$K_{eq} = \frac{[MX]}{[M][X]}$$

Equation 24,

and the mass balance relationships

$$X_{tot} = [X] + [MX]$$

Equation 25

$$M_{tot} = [MX] + [M] = [MX] + \frac{[MX]}{K_{eq}[X]}$$

Equation 26,

where X_{tot} and M_{tot} are the total added concentrations of X and M, a relationship defining [MX] as a function of X_{tot} can be solved from the quadratic root of [MX] in the form

$$0 = ([MX])^2 + [MX]\left(-M_{tot} - X_{tot} - {}^{1}\!/\!_{K_{eq}}\right) + M_{tot}X_{tot}$$

Equation 27.

It follows that:

$$\frac{d[MX]}{dX_{tot}} = \frac{1}{2}\left[\frac{1}{n} - \frac{\frac{X_{tot}}{nM_{tot}} - 1 + \frac{1}{nK_{eq}M_{tot}}}{\sqrt{\frac{X_{tot}^2}{M_{tot}^2} + \left(n + \frac{1}{K_{eq}M_{tot}}\right)^2 - 2\frac{X_{tot}}{M_{tot}}\left(n - \frac{1}{K_{eq}M_{tot}}\right)}}\right]$$

Equation 28

Substituting Equation 28 into Equation 23 allows expression of the molar binding enthalpy in terms of total injectant concentration (X_{tot}):

$$\frac{dQ}{dX_{tot}} = \frac{1}{2}\Delta H^{\circ} V_0\left[1 - \frac{\frac{X_{tot}}{M_{tot}} - n + \frac{1}{K_{eq}M_{tot}}}{\sqrt{\frac{X_{tot}^2}{M_{tot}^2} + \left(n + \frac{1}{K_{eq}M_{tot}}\right)^2 - 2\frac{X_{tot}}{M_{tot}}\left(n - \frac{1}{K_{eq}M_{tot}}\right)}}\right]$$

Equation 29

A more convenient form of Equation 29 is

$$\frac{1}{V_0}\left(\frac{dQ}{dX_{tot}}\right) = \Delta H^{\circ}\left(\frac{1}{2} + \frac{1 - \frac{(1+r)}{2} - \frac{X_r}{2}}{\sqrt{X_r^2 - 2X_r(1-r) + (1+r)^2}}\right)$$

Equation 30

where

$$1/r = c = M_{tot}K_{eq}$$

Equation 31

$$X_r = X_{tot} / M_{tot}$$

Equation 32

Equation 30 is a binding polynomial that describes the molar enthalpy of binding as a function of injectant concentration for a binding reaction with 1:1 stoichiometry under the assumption that the sites are non-interacting. Equation 29 with n representing the number of binding sites, can be used to fit ITC enthalpy data. More complex polynomials are used to describe multiple, interacting and/or sequential binding sites.

From Equation 30, it is apparent that the data fitting by non-linear regression is subject to inherent limitations. A crucial fitting parameter for the accurate determination of the binding constant K_{eq} is the unitless parameter c (Equation 31). The c-value inherently defines the usable experimental ITC window, and optimal fits to the data require c-values of $1 < c < 1000$, or more ideally, $10 < c < 100$.[166]

From the Gibbs free energy relationships

$$\Delta G = -nRT \ln K_{eq}$$

Equation 33

$$\Delta G = \Delta H - T \Delta S$$

Equation 34

it becomes apparent that the determination of both K_{eq} and $\Delta H°$ from a single calorimetric titration provides K_{eq}, ΔH, ΔG, and ΔS for a binding equilibrium. Evaluation

of ΔH as a function of temperature provides ΔCp, yielding a complete thermodynamic characterization of the binding equilibrium.

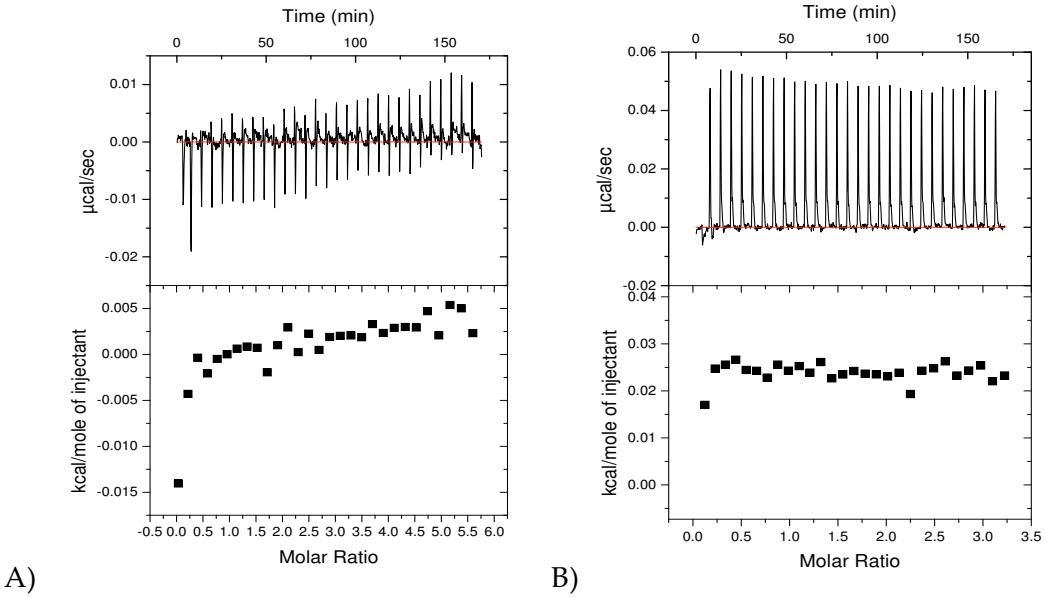


Figure 21: ITC results for CBM3a•cellobiose (A) and cellotriose (B) titrations at 288K

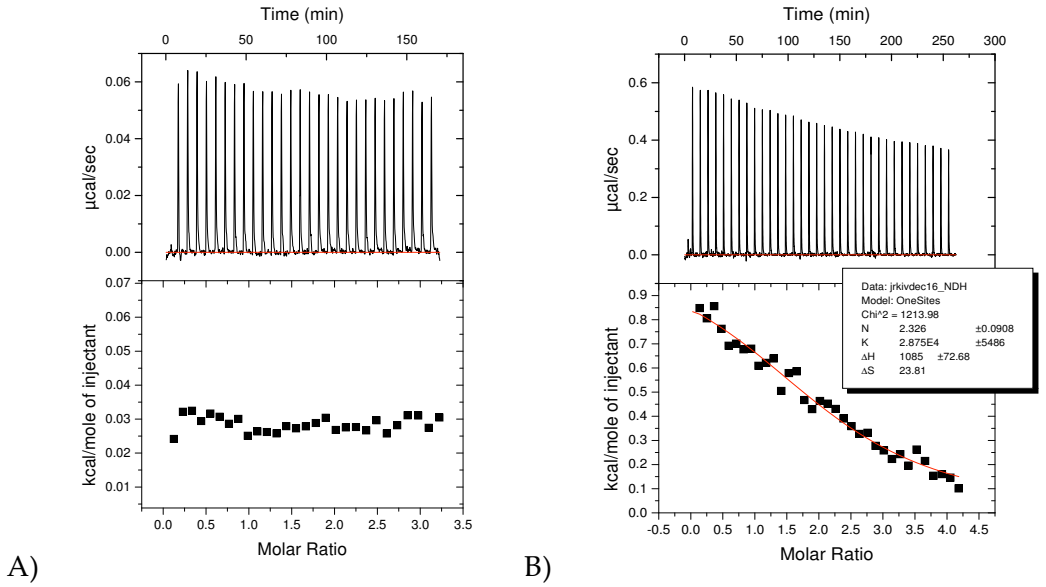


Figure 22: ITC results for CBM3a•cellotetraose (A) and cellopentaose (B) titrations at 288K

The calorimetric titrations between CBM3a and cellodextrins of various sizes are shown in Figure 21 and Figure 22. The titrations depict increased postive heat changes upon injection of the cellodextrin solutions with a general trend towards larger heats with increasing DP. The lack of a binding isotherm for the cellobiose, cellotriose, and tetraose ligands is apparent in the unchanging heat absorption with additional ligand (Figure 21 and Figure 22A). Cellopentaose, however, did appear to yield a decreasing positive enthalpy of binding with increasing mole fraction of injectant, which could be indicative of very weak binding (c-value ≤ 1). Comparison of the binding titrations (i.e. cellodextrin titrated into the CBM3a solution) with reference titrations lacking CBM3a indicated that the observed heats were more likely attributable to ligand dilution than to weak, endothermic binding events. We concluded that CBM3a protein does not bind to water-soluble cellodextrins with a K_a observable by ITC ($\geq 10^3$ M^{-1}).

An important concern with recombinant CBM3a protein (and with all recombinant binding proteins) is its biochemical competency. Unlike enzymes that can be assayed for biochemical competency (i.e. activity) during purification and storage, binding proteins cannot be routinely checked for inactivation. The purification protocol for the recombinant CBM3a begins with cellulose-affinity chromatography; CBM3a eluted from that support is presumably active. In order to confirm the continued binding capacity of the recombinant CBM3a, we titrated a concentrated CBM3a solution into an avicel dispersion (Figure 23). The titration yielded a c-value of 1.7, too low to

provide a meaningful measure of the binding affinity, but the observation of an isotherm did provide a qualitative indication of the binding ability of the protein. It is important to note that the isotherm in Figure 23 was fit to a stoichiometry of n = 1.0 by manually adjusting the concentration of substrate sites on the avicel. This process suggested that the molar density of CBM3a substrate surface sites on avicel was about 2.5 μM/ mg, and provides a physical limitation for the incorporation of sufficient avicel into the VP-ITC instrument in order to deduce meaningful data of the CBM3a-avicel interaction, rendering ITC unsuitable for the characterization of CBM3a binding to insoluble cellulose.

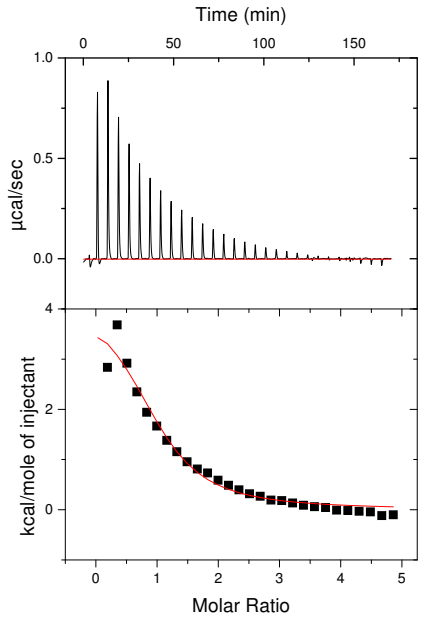


Figure 23: ITC results of CBM3a•avicel titration at 298 K

2.4.2 Low affinity CBM3a-cellodextrin binding tested by ultrafiltration titrations

Given the limitations of ITC to measure CBM3a•cellodextrin binding, we sought other means by which to probe these interactions. To assess CBM3a binding to cellodextrins, we utilized the centricon titration method reported by Menguy et al. for low-affinity binding interactions.[168] In this approach, several samples, each containing an equal concentration of dilute protein, were supplemented with varying concentrations of ligand. The solutions were placed in a Centricon® micro-ultracentrifugation tube and centrifuged at 14 krpm to allow the solution to pass through a semi-permeable membrane that only allows buffer and free ligand to pass. The centrifugation thus concentrates the protein to a final concentration that is close to the K_d of the protein-ligand complex, while simultaneously removing free ligand solution from the formed macromolecular complex.

The free ligand concentration in the flow-through solution was quantified using the BCA reducing sugar assay described in chapter 5 and used to evaluate free ligand concentrations of the concentrate. Plots of ([formed complex]/[total protein]) vs. [free ligand] binding isotherms for cellotetraose and cellopentaose are presented in Figure 24. "Ligand pooling" effects, in which starting solutions of high cellodextrin concentrations (>10 mM) proved too viscous to allow free passage of the ligand solution through the protein concentration membrane, were assessed by blank titration in absence of protein (Figure 24, red squares). Figure 24 illustrates ligand pooling effects, as the data from the

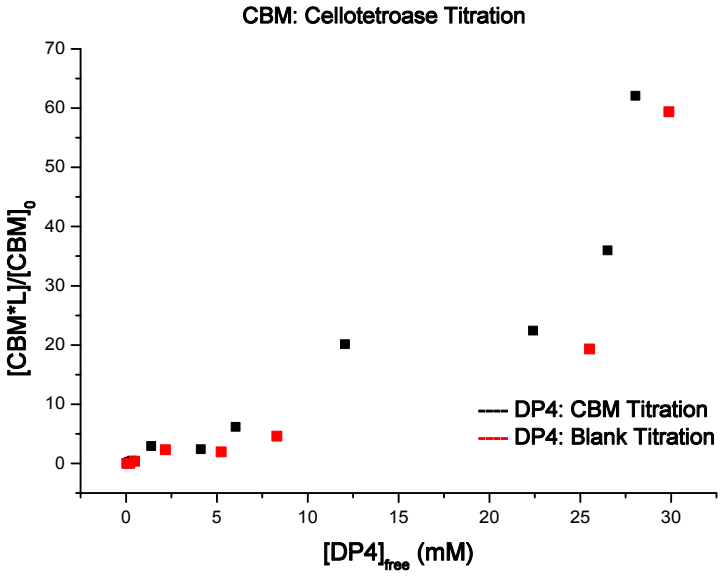
blank titration appears to increase in complex formation at very large ligand concentrations in both titrations with cellotetraose and cellopentaose. In the case of cellotetraose, all of the formed CBM3a•DP4 complex can be accounted for by pooling of the ligand in the concentrate. A slight difference is observable in the CBM3a•DP5 titration, indicating very weak binding between CBM3a and cellopentaose.

The cellopentaose titration was fit to a Michaelis equation using MicroCal Origin-5 non-linear regression analysis software in order to determine an apparent K_d of binding as described by Menguy et al:[168]

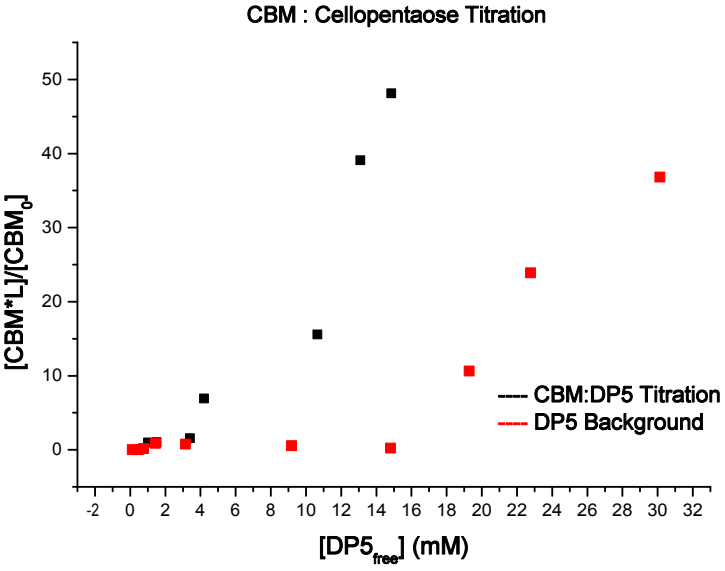
$$\frac{[CBM \bullet DP5]}{[CBM_{Total}]} = \frac{[DP5]}{K_d + [DP5]}$$

Equation 35

The regression produced a dissociation constant of 12 mM. We note a poor correlation in the regression analysis, and centricon titrations produced highly variable stoichiometry values (n). This poor reproducibility in conjunction with the very large dissociation constant (12 mM for cellopentaose, non-observable for cellotetraose) suggests that any observed binding is likely non-specific in nature. In conjunction with the ITC data, these data suggest that the CBM3a protein does not bind small, water-soluble cellodextrins. The implications of this finding are discussed with respect to cellulosomal enzyme targeting effects in the Conclusions section below.



A)



B)

Figure 24: Centricon titration of CBM3a•cellotetraose (A) and cellopentaose (B) binding

2.5 Specific binding at the protein-cellulose interface observed by force spectroscopy[169]*

Cellulose-manipulating organisms use modular cellulose hydrolase enzymes, or *cellulases*, to degrade recalcitrant cellulose microsctructures in order to obtain glucose for respiration.[123, 170] Nearly all cellulophiles incorporate cellulose binding modules (CBMs) and one or more hydrolase domains into fully functional cellulases.[171] Like lipases, cellulose activity cannot be described by classical Michaelis-Menten models of enzyme behavior, since enzyme activity follows *interfacial* binding between the CBM and an insoluble cellulose substrate.[104] The ability to modify traditional hydrolase kinetic models with information regarding specific CBM-cellulose binding is essential for the biophysical characterization of cellulases. Though techniques such as microgravimetry,[156] surface plasmon resonance,[172,173] and UV-visible light spectroscopy[174] are often used to probe the association of CBMs with cellulosic substrates, they are unable to differentiate non-specific adhesion from specific binding. Traditional biophysical methods used to assay protein-carbohydrate interactions, including fluorescence spectroscopy, isothermal titration calorimetry, ultracentrifugation and mass spectrometry are of limited value due to heterogeneity and the insolubility of cellulose substrates.

---

[*] Surface chemistry and force spectroscopy were performed in collaboration with Dr. Carleen M. Bowers.

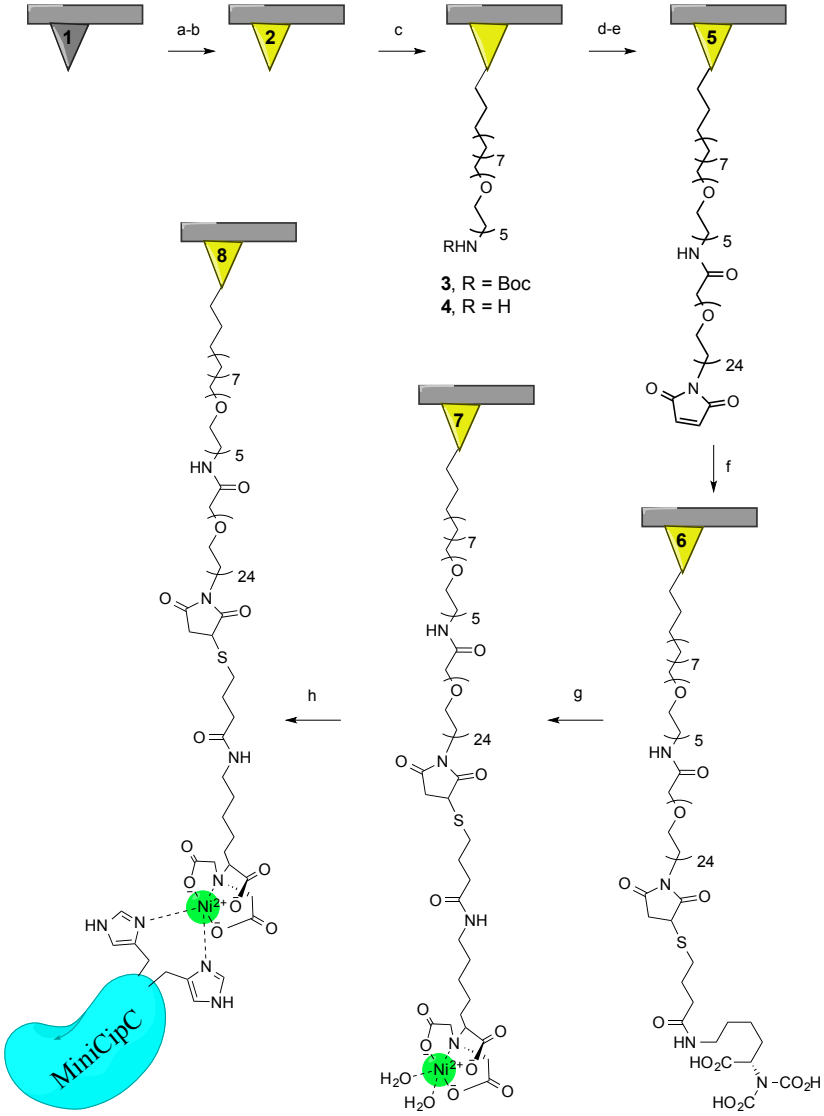
Force spectroscopy, or the forced mechanical unbinding of molecules over nanometer distances using optical tweezers or an atomic force microscope (AFM), offers an intriguing alternative to traditional biophysical tools to observe binding interactions at the solid-liquid interface. Our lab has previously shown that force spectroscopy evalutation of the probability of binding between an immobilized receptor and immobilized ligand in the presence of a free ligand of known binding thermodynamics facilitates quantification of immobilized receptor-ligand binding.[175] Valbuena et al. demonstrated the inherent resistance of Clostridial cellulosome scaffoldin proteins to mechanical stress, a property that makes force spectroscopy an ideal method for measuring Clostridial CBM-cellulose interactions.[56] Xu and co-workers[176,177] recently illustrated the sensitivity of AFM-based recognition force imaging to derive CBM-cellulose unbinding forces (~50 pN at 100 nN/s loading rate) and binding energies of ~15 kcal•mol^{-1} for *Clostridium thermocellum* CBM3a. Like-wise, Kitayoka et al.[178] used AFM to measure binding forces between a hexahistidine (His_6)-tagged CBM from *Cellulomonas fimi*, obtaining a much larger un-binding force of 6.3 nN. The biophysical significance of these reported binding strengths and energies is obscured by the uncertain origin of observed ruptures, which could potentially result from phenomena including protein-cellulose adhesion and force-induced protein denaturation.

We hypothesized that the ability to block non-specific adhesion events with a free ligand in a force spectroscopy experiment would isolate specific CBM-cellulose

binding events from other non-specific interactions. Here, we describe methods to chemically functionalize an AFM cantilever with a Ni(II)-chelating nitrilotriacetate (NTA) group and immobilize a Clostridial CBM3a bearing a His_6 tag. We apply piconewton forces over nanometer distances to disrupt CBM-cellulose binding to a synthetic surface of cellulose nanocrystals (NCs). Finally, we show that we are able to block the CBM-cellulose interaction with a colloidal-suspension of cellulose NCs and observe the return of CBM-cellulose binding upon the removal of the blocking agent.

2.5.1 Surface Chemistry

Immobilization of the CBM3a to an AFM cantilever poses several challenges. In order to reduce non-specific adhesion of protein to the cellulose surface, the protein must be oriented in a favorable binding geometry. We chose to express a recombinant tri-modular *Clostridium cellulolyticum* protein comprised of the CBM3a, a hydrophilic domain, and the first cohesin domain of the cellulosome integrating protein (CipC), following the approach of Pagès et al.[162] This "mini-CipC" (MCC) contains the CBM3a at the N-terminus and an engineered Ni(II)-binding His_6 tag at the C-terminus of the protein. By separating the functional CBM from the His_6 tag via the intermediate hydrophilic and cohesin domains, we ensure that orientation of the CBM3a for producting binding is not precluded by an interfering Ni(II) chelation to the cantilever.



a) NanoStrip, 75 °C, 30 s; b) 5% HF, rt, 1 min; c) 254 nm light at 5 cm, **1**, rt, 30 min; d) 50% TFA in CH_2Cl_2, rt, 30 min; e) 15 mM NHS-dPEG$_{24}$®-Mal, 1% Et_3N in CH_2Cl_2, rt, 2.5 h; f) 10 mM **2**, 5 mM TCEP, 10 mM NaH_2PO_4, pH = 8.1, rt, 13 h; g) 50 mM $NiSO_4$, 20 mM NaHEPES, pH = 7.0, rt, 45 min; h) 40 µM MCC, 10 mM NaH_2PO_4, 140 mM NaCl, 0.05% Tween-20®, pH = 7.4, rt, 30 min.

Scheme 4: Chemical functionalization of AFM cantilever and miniCipC immobilization

The cantilever was modified with a tetracosaethylene glycol (PEG_{24}) linker bearing an NTA moiety for Ni(II) chelation (Scheme 4). We previously reported that the PEG_{24} linker provides sufficient translational and rotational freedom to allow the protein to adopt a productive substrate-binding geometry.[175] Following silicon functionalization chemistry developed in our lab for catalytic microcontact printing[179] and later applied to silicon nitride AFM cantilevers,[180] the cantilever was chemically etched with Nano-Strip® and aqueous HF to form an H-terminated silicon nitride surface. This surface was subsequently alkylated via ultraviolet light-promoted hydrosilylation of the heterobifunctional *tert*-butyl-3,6,9,12,15-pentahexacos-25-enylcarbamate (**1**) linker to yield an *N*-Boc-terminated monolayer. The *N*-Boc functional group was cleaved using trifluoroacetic acid, and the free amine was coupled to a commercial N-hydroxysuccimidyl-PEG-24-maleimide (NHS-$dPEG_{24}$-Mal) linker. A synthetic sulfhydryl-linked NTA molecule (**2**) was used to incorporate the NTA Ni(II) chelator via Michael addition to the thiol-reactive maleimide monolayer. The changing elemental composition at each reaction step was verified by observation of the ratio of carbon to silicon using X-ray photoelectron spectroscopy (XPS).

2.5.1.1 XPS analysis of surface functionalization.

XPS was used to determine the elemental composition during the course of surface functionalization (Figure 25-Figure 29). In each case an element sweep from 0 – 1200 eV was performed to identify the electron absorption bands of elements present on the surface. 10 scans were performed in the C 1s and Si 2p regions to quantify the C/Si ratio[†] of each surface. Due to the poor sensitivity of nickel in XPS analysis, metal chelation to the NTA surface 7 was verified with cobalt via the emergence of the Co-2p band using 10 scans in the Co 2p absorption region in XPS. Interestingly, metal chelation was not observed under non-buffered soaking conditions (50 mM $CoCl_2$ or $NiSO_4$) common to metal loading on immobilized metal-affinitiy chromatography resins. This effect is likely due to a surface-induced increase in the pK_a of the NTA moiety as described by Whitesides and co-workers[181] and validated here via goniometric pH titration (surface **5**, Table 5).

[†] The measured C/Si ratio is normalized by multiplying by the Si/C relative sensitivity factor ratio of 1.1799 that is specific to this instrument.

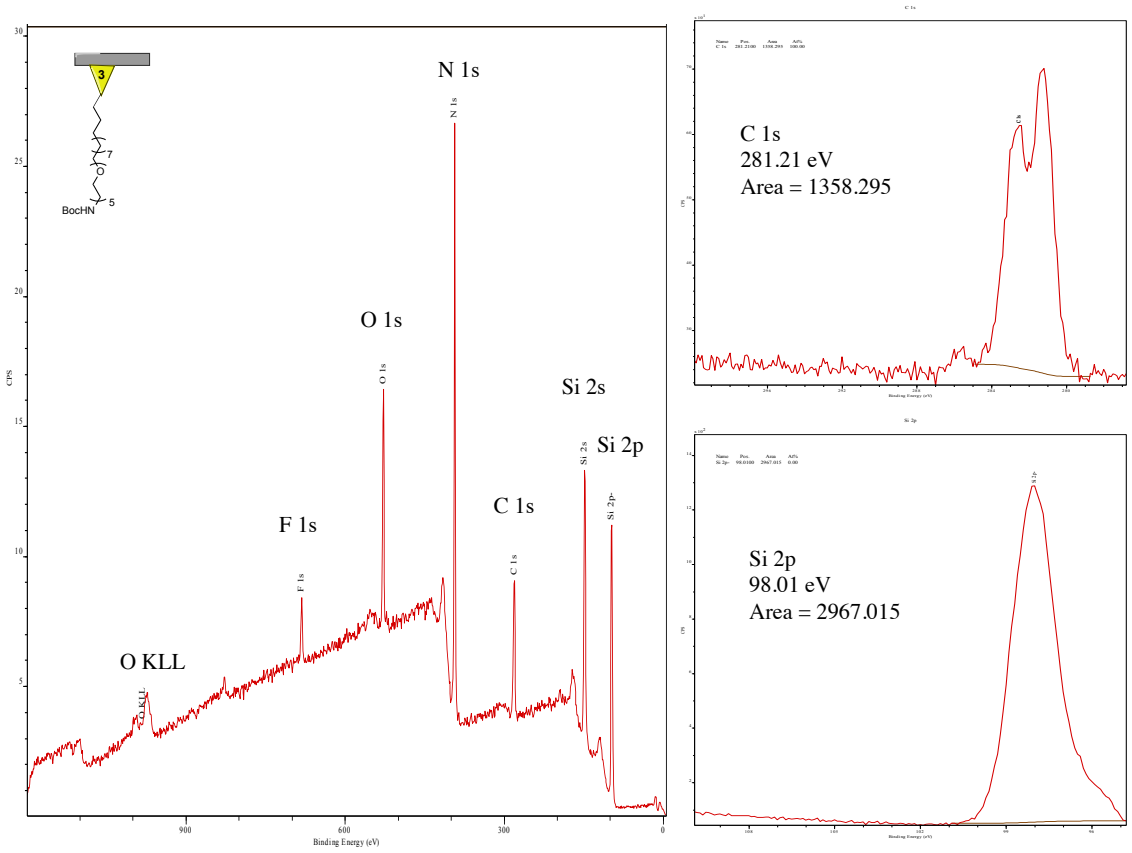


Figure 25: XPS data for R-NHBoc surface 3

Surface 3 XPS data indicates presence of C, N, O, Si, and a slight contamination of F.

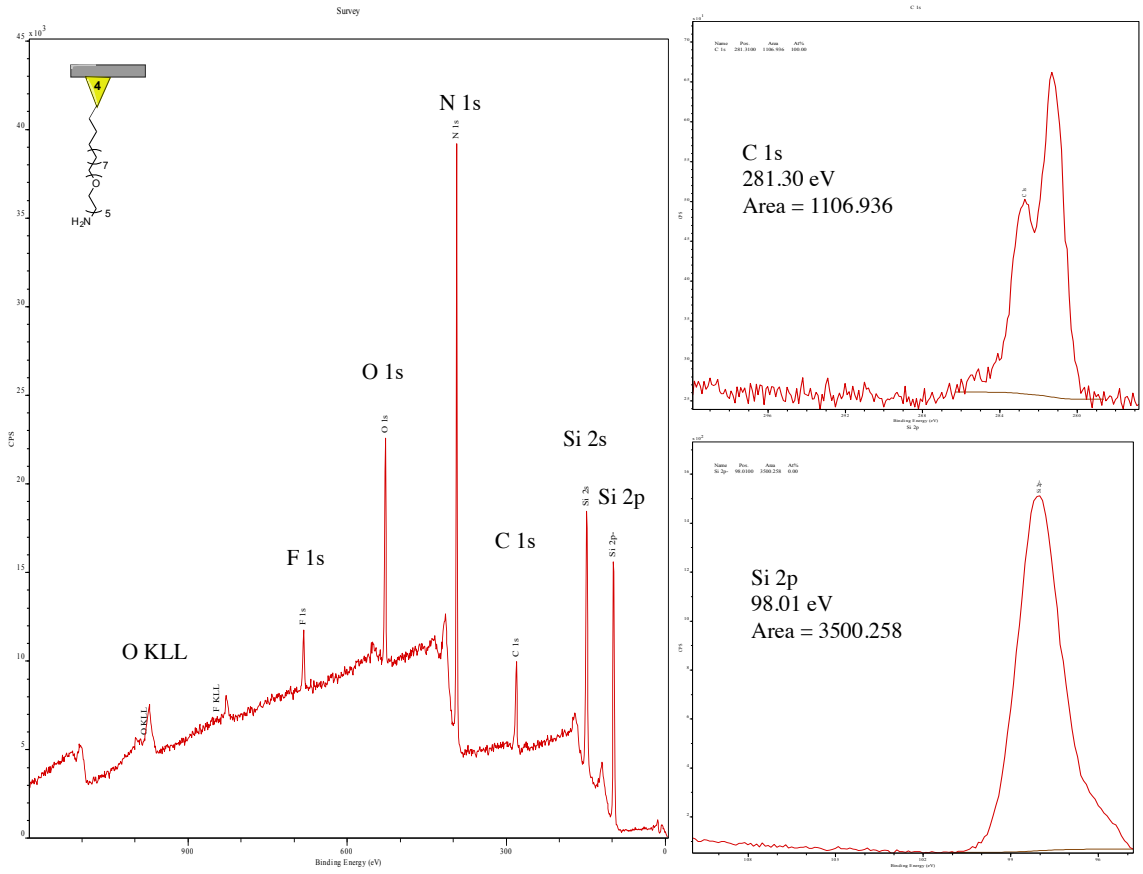


Figure 26: XPS data for R-NH$_2$ surface 4

Surface 4 XPS data indicates presence of C, N, O, Si, and a slight contamination of F. A slight decrease in the C/Si content with respect to surface 3 indicates removal of the Boc group.

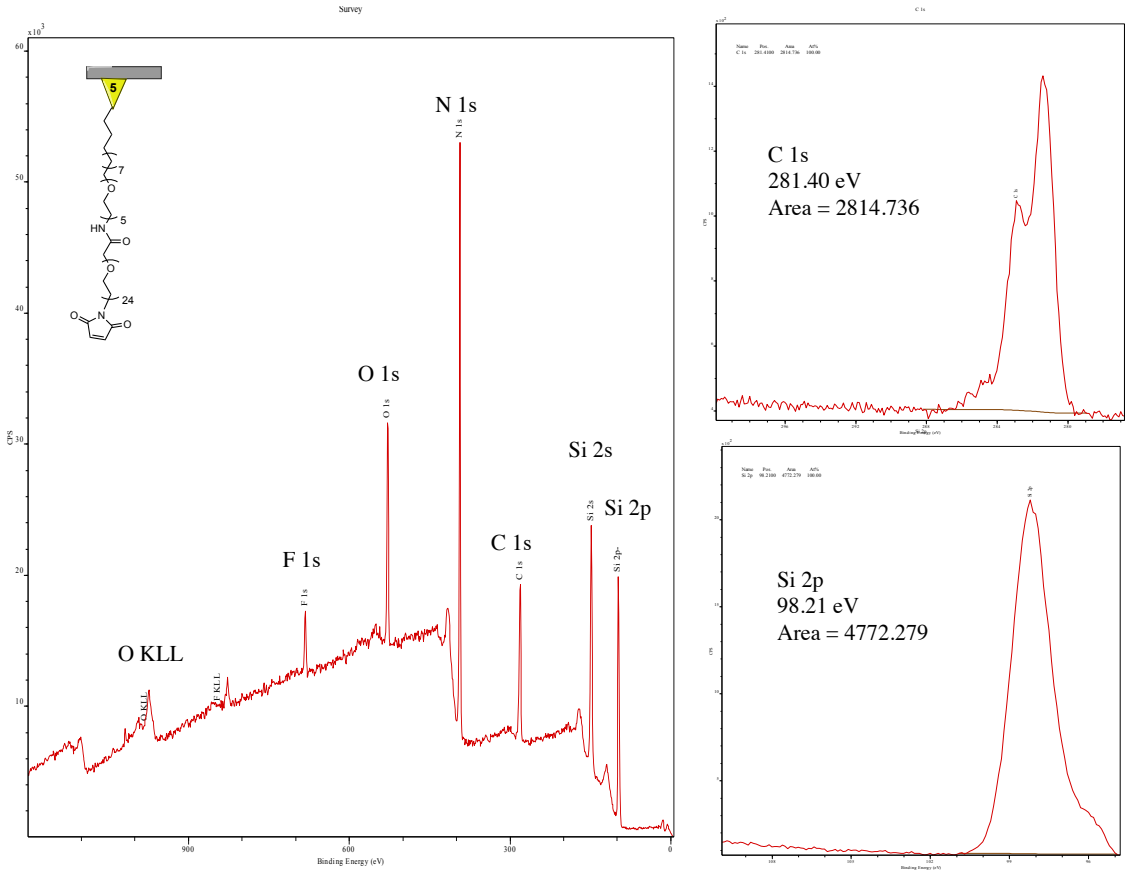


Figure 27: XPS data for R-PEG$_{24}$-Mal surface 5

Surface 5 XPS data indicates presence of C, N, O, Si, and a slight contamination of F. An increase in the C/Si content with respect to surface 4 indicates addition of the PEG-ylated linker.

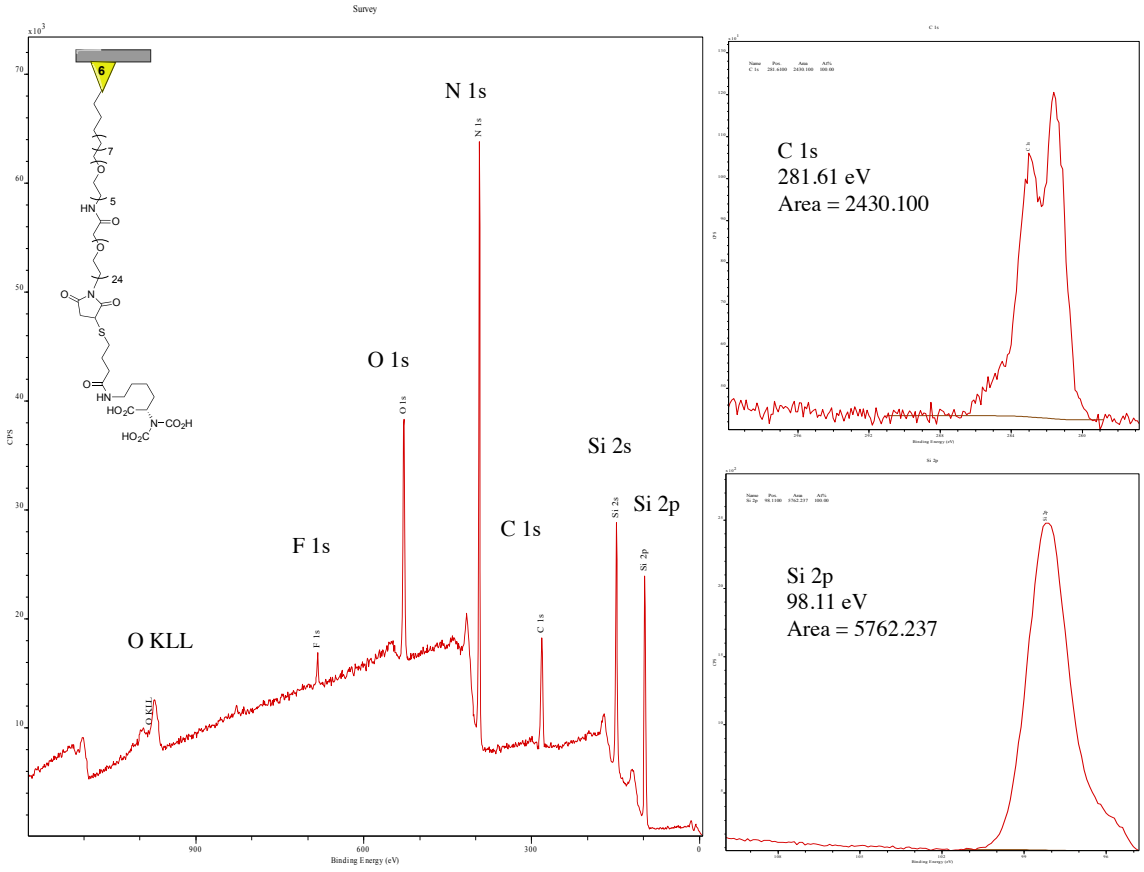


Figure 28: XPS data for R-NTA surface 6

Surface 6 XPS data depicts C, N, O, Si, and minimal F. The C/Si ratio is no longer quantitative due to the disordered nature of the mixed monolayer at this stage of functionalization. The disorder is also evident in the large hysteresis from goniometric analysis (Table 5).

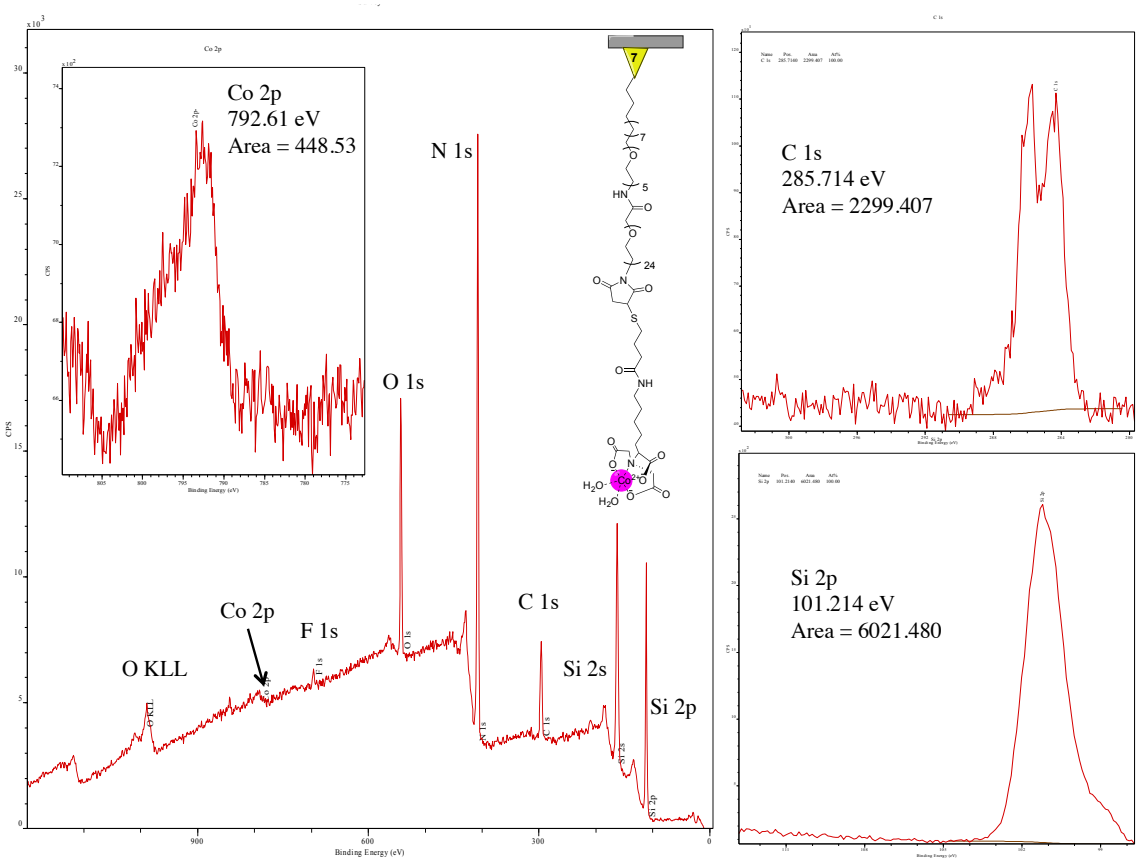


Figure 29: XPS data for R-NTA-Co surface 7

XPS data from surface 7 depicts C, N, O, Si, minimal F, and (most importantly) Co. The C/Si ratio is no longer quantitative due to the disordered nature of the mixed monolayer at this stage of functionalization. The presence of the Co-2p band (793 eV) at pH = 6.8 supports the presence of Co-chelating NTA moieties on the surface.

2.5.1.2 Surface characteristics revealed by goniometric analysis of water contact angles

Changes in chemical composition of the surface at each reaction step alter the hydrophobicity of the surface, an effect evident in differing water contact angles after

each reaction (Table 5). A clear decrease in contact angle with increasing pH is consistent with greater surface wetting as the NTA moiety is ionized. Although the reported pK_a of NTA is ~3 (pK_a = 3.03, 3.07, 10), the surface appears to perturb the pK_a upward to near 5.5. As a result, buffered $NiSO_4$ at pH > 6.5 was necessary to efficiently immobilize Ni(II) to the surface.

Table 5: Surface functionalization monitored by XPS, water contact angles, and goniometric pH titration

Surface*	C 1s/ Si 2p (peak area)	S 2p	Co 2p (position, eV)	Adv. Contact Angle (°)	Rec. Contact Angle (°)	pH
3, R-NHBoc	0.54	ND§	ND	67	58	ND
4, R-NH_2	0.39	ND	ND	49	38	ND
5, R-PEG-Mal	0.71	ND	ND	60	46	ND
6, R-Mal-NTA	0.61	NO⌘	NO	49	32	2.0
				52	35	3.1
				51	28	4.0
				53	27	5.1
				45	20	6.1
				45	20	7.0
				46	17	8.0
				43	16	9.1
				42	12	10.1
				42	10	11.1
7, R-NTA-Co	0.64	NO	NO	ND	ND	5.5
		NO	792.6	ND	ND	6.8
		NO	793.5	ND	ND	7.4

*Surfaces from **Error! Reference source not found.**; §ND: data was not determined; ⌘NO: peak was not observed.

2.5.1.3 Model cellulose I surface characterized by AFM

Cellulose nanocrystals (NCs) were spun on silicon into a 24.5 nm film using the method of Edgar and Gray[182] as described by Wågberg et al.[183] AFM images confirmed that the cellulose NCs were uniformly spread across the substrate surface (Ra = 1.5, Rq = 1.8), allowing effective CBM3a-cellulose binding at any location on the wafer (Figure 30). Inspection of the images with NanoScope software indicated the NCs were typically 25 – 30 nm in width, 80 – 100 nm in length, and 8 nm in depth. Unlike other cellulose films, cellulose NCs can be prepared in a mild, aqueous buffer that is compatible with cellulose binding proteins. This unique feature enabled the use of a colloidal suspension of cellulose NCs to block specific binding interactions between the cantilever-bound CBM3a and the cellulose surface, thus differentiating specific interfacial binding events from non-specific protein-cellulose adhesion (*vide infra*).

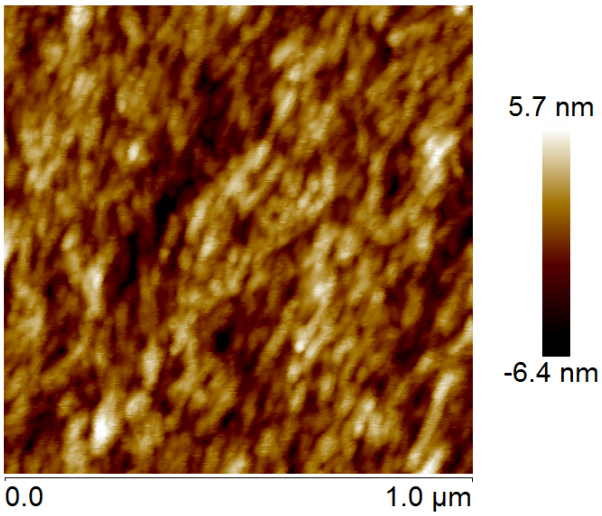


Figure 30: AFM image of spun-cellulose I nanocrystals

2.5.2 Force spectroscopy controls and MCC-cellulose binding

AFM is a sensitive tool, measuring cantilever tension at pN loads and stage retraction over nm distances. As a result, the analysis of force spectroscopy data is subject to errors associated with surface vibrations and tip fluctuations that can alter rupture morphology and lead to inconsistent identification of ruptures. Other experimental parameters, such as contact force and dwell time, also alter rupture morphologies and impact the probability of observing a rupture even within a single data set. To minimize these concerns, we rely on the guidelines outlined by Bowers et al[184] for objective analysis of rupture data.

2.5.2.1 NTA-Ni(II)-Histag forced rupture profile

A prerequisite to the use of AFM to characterize CBM3a-cellulose binding is an assessment of the force and distance limitations of the molecular system. The CBM3a-cellulose interaction is dominated by non-covalent hydrophobic contacts.[153] To assess the mechanical strength of these interactions in comparison to the non-covalent coordination of the His_6 tag and the NTA-Ni(II) complex, a model peptide with the primary sequence CGWGGHHHHHH was synthesized using microwave-assisted solid-phase peptide synthesis. The peptide cysteine residue was used to covalently link the peptide to a maleimide-terminated silicon surface using chemistry similar to that described above. A Ni(II)-loaded NTA monolayer on the AFM cantilever was gradually brought into contact with the peptide surface. Upon retraction, the deflection of the

cantilever yielded a force vs. distance plot characteristic of force-induced rupture of the His_6-Ni(II)-NTA interaction. We repeated the process for 250 cycles and plotted the frequency of observed rupture forces and lengths (Figure 31). The frequency histograms were fit to an inverse Gaussian distribution function to determine statistically significant forces and lengths. The mean Ni-dependent rupture force was 121 ± 7 pN at a retraction velocity of 200 nm•s^{-1},* and the mean rupture length was 15.5 ± 0.4 nm, values that agree well with a molecular system length of about 13 nm (3 nm peptide + 10 nm PEG_{24}).

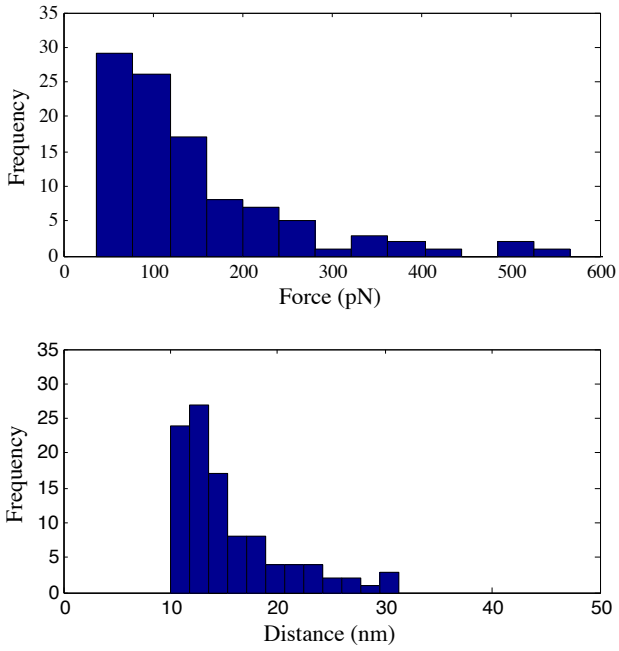


Figure 31: Histag rupture force and length histograms

* The retraction velocity corresponds to loading rate of ~0.4 nN•s^{-1} as calculated for the presented molecular system by Bowers et al. (2013).

2.5.2.2 Cellulose-tip and cellulose-miniCipC interactions

To determine the inherent adhesive forces associated with cantilever-cellulose contact, we plotted the systematic rupture frequencies at varying distances for a "blank" NTA-terminated cantilever in contact with a cellulose film. The probability of observing a rupture event (P_{bind}) was 14% over 150 pulls, indicating minimal adhesion between the tip and cellulose surface. The adhesion forces observed varied widely, though most were below 150 pN. The rupture lengths distribution indicated no statistically significant mean rupture length, consistent with non-specific adhesion.

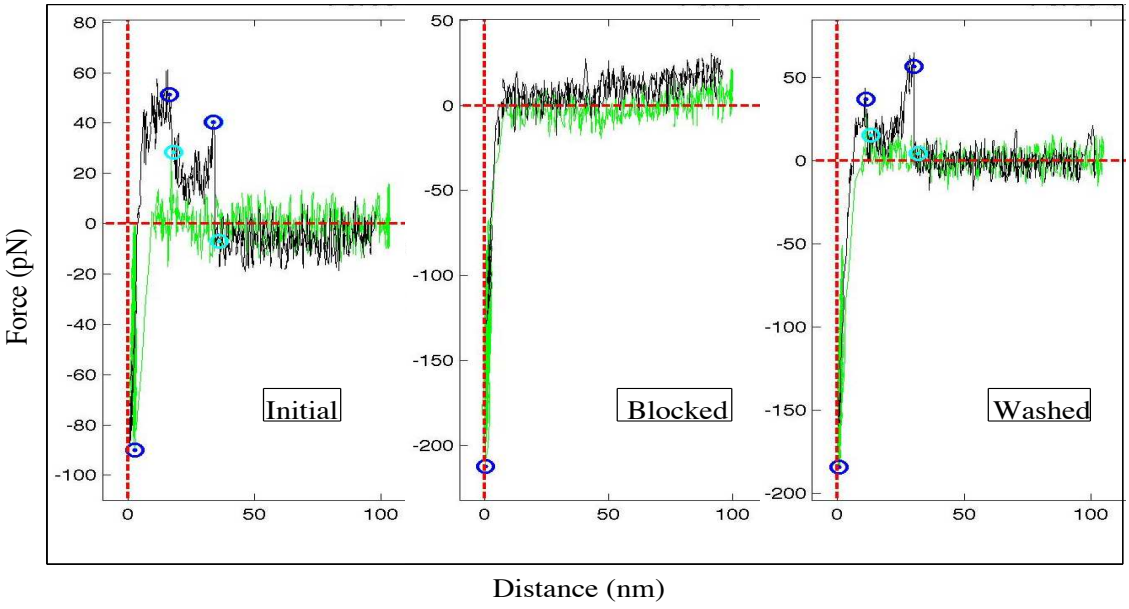


Figure 32: MiniCipC-cellulose forced rupture curves

With a firm understanding of the system characteristics, we next sought to characterize specific (e.g. "blockable") CBM3a-cellulose interactions. Initially, the cantilever-immobilized MCC was brought into gentle contact with a nanocrystalline

cellulose surface, and the surface was retracted to generate a force vs. distance plot (Figure 32). Next, a blocking solution of cellulose NCs in buffered water was introduced into the AFM flow cell. Initial attempts to monitor binding events in the presence of the NC suspension were unsuccessful due to large background cantilever movement, presumably attributable to the presence of the NCs. Fortunately, the CBM3a-cellulose interaction is highly salt dependent,[71] and removal of the NCs with buffered saline left intact the bound MCC-NC complexes but successfully removed unbound NCs and the large NC-induced noise. The NCs were removed from the MCC molecules with pure water—a process routinely used in the purification of CBM3a via cellulose-affinity chromatography[71, 162], and the MCC-cellulose binding returned in the presence of buffered saline, albeit with a large reduction in total binding events as compared to the initial data set (Figure 33).

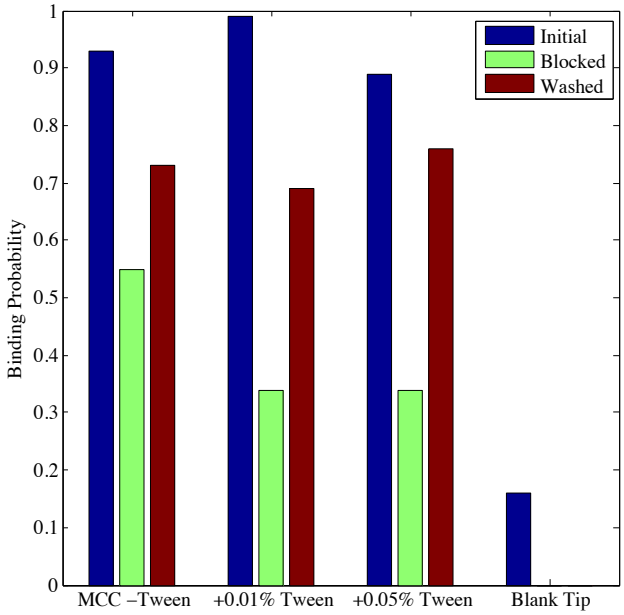


Figure 33: MiniCipC-cellulose binding probabilities

Each cycle was repeated in excess of 100 pulls to yield frequency histograms of rupture forces and lengths. Significantly, a reduction in P_{bind} from 93% to 55% was observed from the initial to blocked binding experiments. The binding probability increased (73%) following removal of the blocking agent, indicating a significant number of specific, "blockable" interactions. The large binding probability in the presence of blocking agent is consistent with extensive non-specific adhesion. Gratifyingly, the highest frequency of blockable rupture events was observed at a rupture length of 32 ± 1 nm (Appendix A), a distance consistent with a predicted molecular system of about 30 nm (20 nm protein + 10 nm PEG-linker). The distribution of rupture lengths may have been skewed by the presence of short non-specific tip-cellulose adhesion events (*vide supra*). The mean rupture force of 50 ± 1 pN is slightly larger than the 33.3 pN rupture

force reported by Xu et al[177] at 16 nN/s and is well below the 120 pN binding threshold reported here for the His_6-Ni(II)-NTA interaction.

A significant reduction in rupture events was also apparent in the force and length histograms after the initial stage of the experiment. This drop in binding frequency upon blocking and washing is also consistent with substantial levels of non-specific adhesion in the initial stage of the experiment. Because we cannot rule out elevation of the mean rupture force and diminution of rupture lengths as a result of background adhesion between MCC and the cellulose surface, we sought to remove non-specific contacts between the AFM tip and cellulose surface using buffer additives. The experiments were repeated with a non-ionic detergent (Tween-20®), which should disrupt weak hydrophobic contacts between the PEG_{24} linker and protein with the cellulose surface. As shown in Figure 32, the addition of up to 0.05% Tween-20® resulted in a slightly lower initial P_{bind} of 89% but yielded a satisfactory blocking efficiency (P_{bind}=34%) and a substantial return of binding upon removal of the blocking agent (P_{bind}=76%).

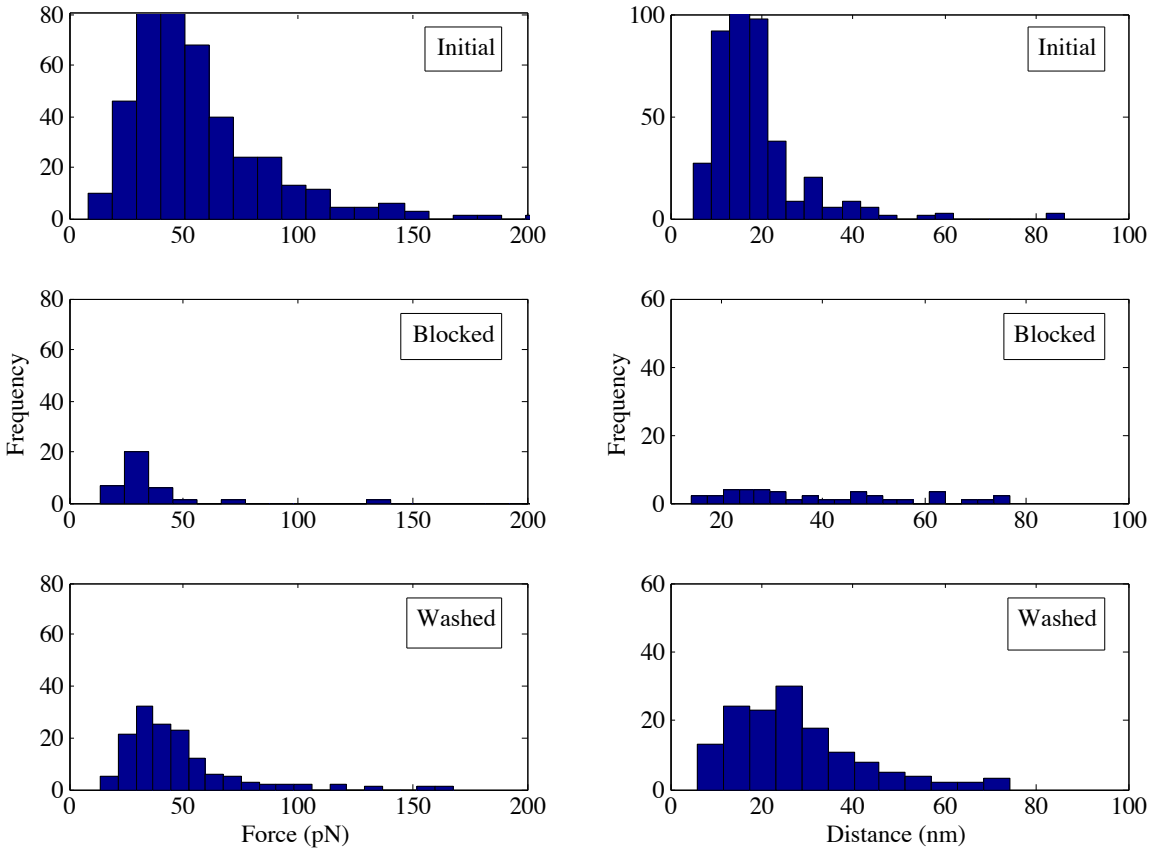


Figure 34: MiniCipC-cellulose rupture force and length histograms

MCC-cellulose rupture force and length profiles are presented in Figure 34. We observed an initial mean rupture force of 52 ± 1 pN which agrees well with the initial force in the absence of Tween-20®; however upon blocking the tip with a NC suspension and subsequently removing the blocking agent in pure water, that mean shifted to a smaller rupture force of 45 ± 2 pN, closer to the values of Xu et. al.[177] The initial mean rupture length was 17.5 ± 0.3 nm, shifting to a longer mean of 28 ± 1 nm after blocking

agent was removed.[†] The shift to lower rupture forces and longer rupture lengths upon addition and removal of the blocking agent suggests a different chemical environment exists in the initial and washed stages of the experiment. We hypothesized that the cellulose blocking agent *reversibly* inhibits specific CBM3a binding sites and *irreversibly* disrupts non-specific protein-cellulose interactions.

2.5.3 Galectin-3 force spectroscopy: elucidating specific ruptures

To probe this hypothesis, we exchanged MCC with the similarly sized lectin, murine galectin-3 (G3, 35 kDa). We have previously detailed the specific affinity (K_a ~ 6400 M^{-1}) of G3 for small galactose-based carbohydrates and demonstrated its suitability for force spectroscopy experiments.[175, 180, 184, 185] To our knowledge, G3 has no reported affinity for polysaccharides or cellulose and should serve as an appropriate control to identify non-specific protein-cellulose interactions in force spectroscopy and to test our blocking methodology. If the cellulose NC blocking agent does irreversibly inhibit non-specific protein-cellulose interactions, we should initially observe G3-cellulose

---

[†] Deviations in the rupture lengths towards smaller distances is expected when considering the cone shape of the cantilever tip. Few molecules are located near the edge of the tip, where the rupture distance is likely to demonstrate the molecular length of the system.

interactions blockable by NC suspension, but binding should not reappear after washing.

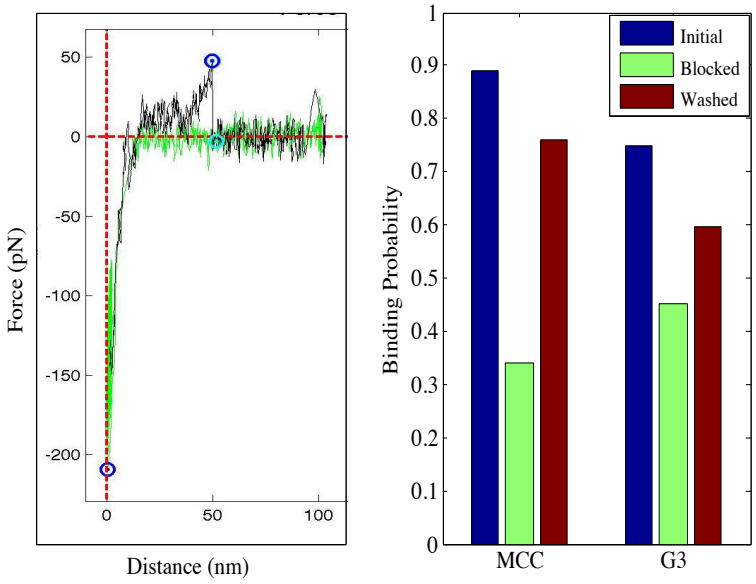


Figure 35: Cellulose-Galectin 3 unbinding curve and binding probability

As with MCC experiments, G3 was immobilized to the AFM cantilever via a NTA-Ni(II)-mediated His_6 tag ligation of the protein. The tip was brought into contact with cellulose for 1 second, and the surface was removed at a retraction velocity of 200 nm/s. The process was repeated for 250 pulls over three cycles in the absence (initial/ washed) or presence (blocked) of the cellulose NC suspension. The observed rupture morphologies for G3-cellulose interactions were nearly identical to those recorded in our MCC-cellulose experiments and the G3-lactose force curves reported by our previous studies.[175, 180, 184] A representative rupture curve and P_{binds} are given in Figure 35. The change in P_{bind} with and without the cellulose blocking agent qualitatively suggests blockable interactions were observed as the initial P_{bind} of 75% reduced to 45% under the

blocked conditions. Interestingly, P_{bind} increased to 60% in the washed experiment, suggesting the G3-cellulose interactions were reversible and specific; however, the differences in P_{bind} are too small to quantitatively determine if G3 reversibly binds cellulose or if the increased P_{bind} in the washed experiment is of an alternative origin.

To resolve the differential P_{binds} between the blocked and washed G3-cellulose binding experiments, we examined the change in rupture force and length distributions under initial, blocked, and washed conditions (Figure 36). We observed little variation in the force distributions between different pulling cycles. Strikingly, the forces observed for G3-cellulose un-binding (~46 pN) were nearly identical to the MCC-cellulose rupture forces observed in the regenerated binding (washed) experiments (45 pN). Though this could suggest that G3 interacts with cellulose in a fashion analogous to MCC, the similarity in rupture forces more likely demonstrates that non-specific protein-cellulose contacts are mechanically similar to specific CBM3a-cellulose interactions. This observation is not entirely unexpected, as qualitative observations of a highly salt-dependent binding between two very flat cognate CBM and cellulose binding interfaces clearly illustrate the possibility of a mechanically weak (i.e. low force barrier to unbinding), yet thermally stable (i.e. large thermal barrier to unbinding), binding driven predominantly by hydrophobic desolvation.[71, 153]

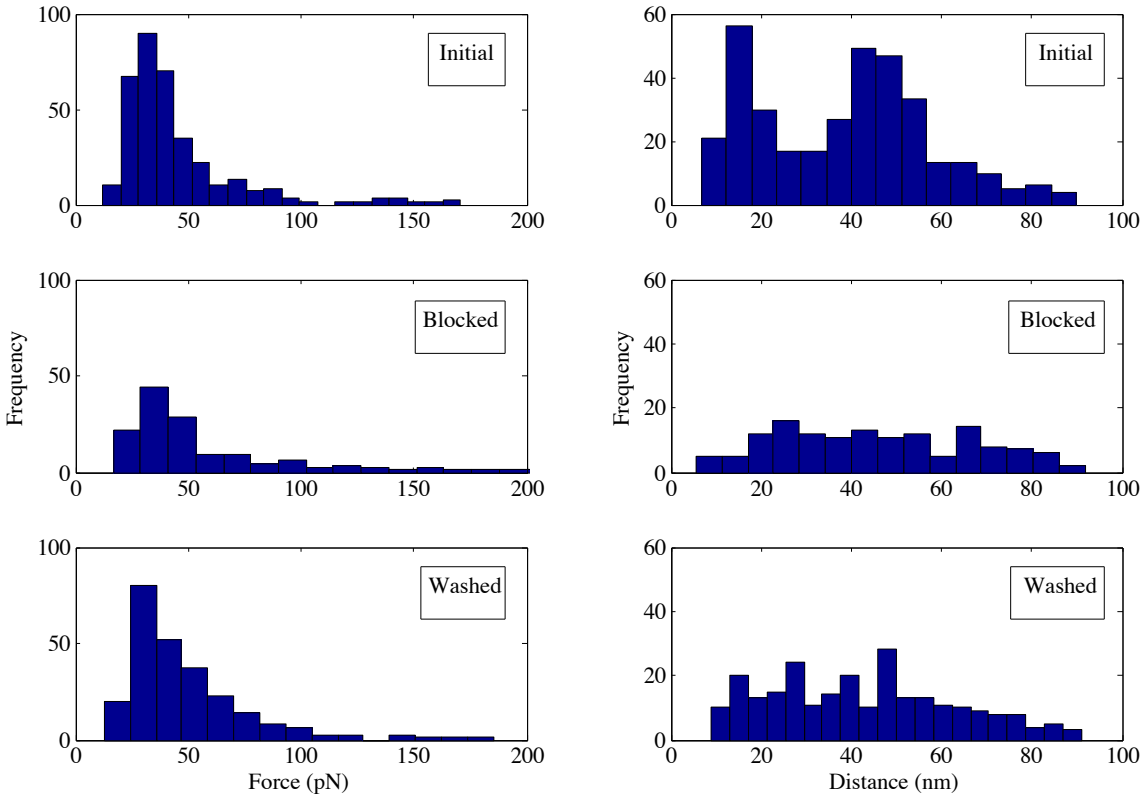


Figure 36: Cellulose-Galectin 3 rupture force and length histograms

As we were unable to use force distributions to differentiate specific from non-specific binding, we analyzed rupture length distributions to identify qualitative differences in MCC-cellulose and G3-cellulose ruptures (Figure 36, *right*). In the initial G3-cellulose un-binding experiment, we observed two classes of rupture lengths characterized by short (17 nm) or long (50 nm) lengths. As with MCC, we attribute the two length classes to be largely due to short linker/tip-cellulose interactions and longer protein-cellulose interactions. Under the conditions of the blocked experiment, each class of ruptures is ablated, and neither the short nor long class of ruptures appears to regenerate after washing away blocking agent. Indeed, the cellulose NC suspension

appears to irreversibly block both the short and long classes of G3 ruptures, while reversibly ablating non-specific contacts, apparent in the increased P_{bind} in the washed (60%) as compared to blocked (45%) experiment across nearly all measured distances. When the G3-cellulose rupture lengths are interpreted in comparison to MCC-cellulose rupture lengths, we clearly observe reversibly blockable length distributions in the expected regime for MCC-cellulose interactions (25 – 40 nm), although we failed to regenerate the initially observed G3-cellulose interactions in the length range for G3-cellulose contacts (40 – 60 nm).

2.6 Conclusions

We have reported the production of cellodextrin mixtures using a combination of acid hydrolysis of avicel and pivaloylysis of cellulose triacetate. The mixtures were subsequently purified via organic-phase separation over octadecyl silica with RP-HPLC or tandem calcium affinity and size exclusion chromatographies in aqueous milieu. Clostridial CBM3a domains were heterologously expressed in *E. coli* and purified by cellulose affinity chromatography, size exclusion, and IMAC methods. Subsequent calorimetric evaluation of the CBM3a•cellodextrin bindng provided no discernible binding affinity for the investigated binding partners. Additionally, centricon ultracentrifugation titrations for evaluating low affinity binding further supported the lack of association between CBM3a and water-soluble cellodextrins.

CBM3a•cellulose interfacial binding was explored using force spectroscopy. A generally useful procedure was described for the stepwise chemical modification of silicon nitride AFM cantilevers with hydrosilylation chemistry and orthogonal chemical ligation protocols to instill NTA moieties for the Ni(II)-mediated immobilization of His_6-tagged proteins. Frequency analysis of the forced unbinding profile of an immobilized Clostridial CBM3a-containing miniCipC protein opposite a nanocrystalline cellulose surface clearly demonstrated the mechanical rupture of specific MCC-cellulose interactions under a force load of 45 pN at 0.4 nN/s over a distance of 28 nm. Specific binding could be disrupted with a suspension of cellulose nanocrystals and regenerated by the removal of the NCs with pure water. Upon repeating the experiment with murine galectin-3, which does not bind cellulose, nearly identical force-failure profiles were observed as compared to MCC. However, upon evaluation of the frequencies of rupture lengths, distinct rupture length distributions describing G3-specifc ruptures were not regenerated in the washed stage of the experiment, indicating that no such specific interactions existed between G3 and cellulose. On the other hand, MCC-cellulose rupture length distributions could clearly be regenerated in the washed stage of the MCC-cellulose binding experiment. These data illustrate how rupture length profiles and blocking agents must be used together to differentiate specific (eg. reversible and "blockable") binding events versus nonspecific (eg. irreversible and/or un-blockable) adhesion. Furthermore, our results offer the first clear indication of specific binding at

the CBM3a-cellulose interface by way of accounting for non-specific protein-cellulose contacts in the use of galectin-3-based negative control.

Given the large discrepancy in reported un-binding forces for CBM-cellulose association (51 pN at 100 nN/s[177] vs. 6.3 nN[178]) and the evaluation that force-derived CBM-cellulose binding energies (~14 kcal•mol^{-1})[177] exceed protein folding energies (~7 kcal•mol^{-1})[186], our results provide a cautionary tale for the determination of binding energies via the interpretation of binding forces alone. As such, we suggest that evaluation of changing binding probabilities during a competitive binding titration[175] is the preferred method for determining immobilized protein-substrate binding energies and binding constants using AFM.

We set out to probe the nature of CBM3a-cellulose association and enzyme targeting in cellulosome action. While our results are not dispositive with respect to this question, they do offer important insight into the cellulose-binding mode of *C. cellulolyticum* CBM3a, and suggest a novel and previously unknown binding mode. Our results clearly demonstrate that CBM3a specifically binds crystalline cellulose in reversibly blockable and a salt-dependent manner, but does not bind soluble linear oligosaccharides, at least through the pentaose. Our ITC results of the CBM3a•avicel binding (Figure 23) suggest the binding is entropically driven at 298 K, suggesting a role of hydrophobic desolvation in binding. Although a co-crystal structure has not been reported, an examination of the binding face of CBM3a offers clues as to how such a

binding might occur. The lower beta sheet of the protein, containing the canonical sugar-binding residues (Figure 37; S14, S19, Y23, W58, H61, Y70, W123, and Y140), is ideally structured to interact with a perimeter ring of saccharide residues on adjacent chains and encompassing the large, flat face of a cellulose sheet; such a binding would presumably be driven largely by hydrophobic desolvation. Such a binding mode would also rationalize the lack of binding of even large linear oligosaccharides, since critical binding contacts require adjacent polymer chains. Rather, the proposed mode would exploit the unique crystalline structure of cellulose, and offer a powerful mechanism of enforcing specificity – i.e. targeting – in a milieu contaminated with myriad soluble substrates of identical linear (or primary) structure.

This proposed binding mode is also consistent with the O-ring theory of Bogan and Thorn[187] in which a ring of residues at the binding interface of the protein occludes water from a "hot spot" residue at the center of the protein binding site. Hot spots, first described by Clackson and Wells,[188] are protein residues that account for the majority of the binding free energy in protein-protein interactions. Application of the O-ring theory to CBM3a-crystalline cellulose binding provides a molecular rationale for the apparent specificity of CBM3a for very flat carbohydrate surfaces, and lack of specificity to soluble cellodextrins. The implications of this rationale provide insight into enzyme targeting effects proposed by Fierobe and co-workers.[121, 122, 131] Thus, we suggest that future work to evaluate targeting effects in the cellulosome should focus on identifying potential hot

spots for affinity on the CBM3a surface and also evaluate the role of cellulose crystallinity for CBM3a-cellulose association.

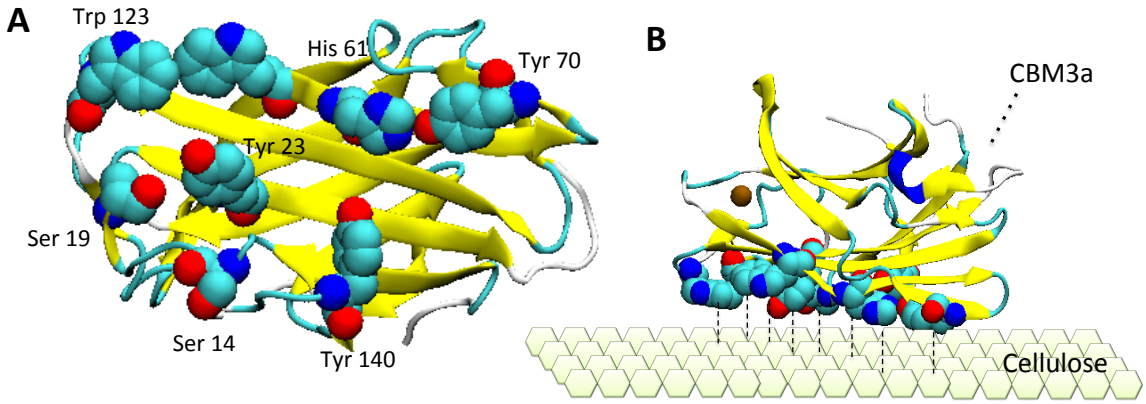


Figure 37: *C. cellulolyticum* CBM3a X-ray crystal structure (A; 1g43) and proposed binding mode to cellulose (B)

3. Comparison of cellulase activity assays and synthesis of a flourogenic cellohexaoside

3.1 Introduction

The functional study of cellulases is a broad field limited by a dearth of methods for quantifying cellulase activity. The challenges posed by cellulase assays are not insignificant. Natural cellulose fibers are a mixture of highly recalcitrant, insoluble, crystalline cellulose I microfibrils containing disordered regions of amorphous cellulose encapsulated in layers of hemicellulose, pectin, and lignin; cellulase accessibility is limited without chemo-enzymatic pretreatment of the natural substrate to expose the cellulose substrate. Pure cellulose is insoluble, has no absorptions in the UV/vis range, and chemically inert—characteristics that constitute a poor substrate for routine enzyme assays. The field of cellulase study has broadly relied on two classes of assay to quantify cellulase activity: i) the discontinuous analysis of soluble sugar content of the reaction mixture, and ii) the implementation of complex and specialized technologies to follow cellulase activity via real-time, direct measures of cellulose depolymerization. These classes of assay are described below.

3.1.1 Discontinuous cellulase assays

3.1.1.1 Chemical reducing sugar assays

The current IUPAC method to monitor cellulase activity utilizes 3,5-dinitrosalicylic acid (DNS) to oxidize soluble sugar products from cellulase-catalyzed cellulose depolymerization, namely cellobiose and glucose.[189] In addition to serving as

an indirect indication of cellulase activity, the method is discontinuous, laborious, and insensitive below micromolar concentrations. Other oxidants and dyes, such as ferricyanide,[190] 4-hydroxybenzhydrazide (p-HBH), bicinchoninnic acid (BCA),[191] tetrazolium blue, and 3-methyl-2-benzothiazolinonehydrazone (MBTH) have been used to increase sensitivity;[189] however, many of these dyes react with proteins, and all rely on the indirect and discontinuous indication of cellulase activity under enzyme-denaturing conditions.

Table 6: Comparison of reducing sugar assays for CelF avicellase quantitation

Substrate	Oxidant	Conditions	CelF activity*	Comments
Avicel	K_3FeCN_6	Alkaline cyanide and sodium dodecyl sulfate in boiling water for 15 min.	20 – 50 mg•$L^{-1}hr^{-1}$	Fe reagent is not compatible with protein and use of SDS and boiling water limit use in high-throughput or continuous cellulase assay.
Avicel	DNS	1% DNS, 0.2% phenol, and 1% buffered NaOH in boiling water for 5 min.	Not detectable.	The reagent was not sensitive enough to measure crude CelF activity. The reaction is run under enzyme-denaturing conditions.
Avicel	p-HBH	1% pHBH in 1M NaOH at 75 °C for 30 min.	29 mg•$L^{-1}hr^{-1}$	pHBH is insensitive to protein and only requires a reaction temperature of 75 °C; though alkalinity is not compatible with enzyme.
Avicel	BCA	0.2% BCA + 0.1% Cu^{2+}, pH ~9 at 70 °C, 30 min.	~50 mg•$L^{-1}hr^{-1}$	Cu-BCA cross-reacts with protein.

*CelF concentrations typically in the low micromolar range from crude preparations.

In our own search for a sensitive, facile assay of cellulase activity, reducing sugar produced from *C. cellulolyticum* CelF-catalyzed hydrolysis of microcrystalline cellulose (avicel) was measured under several conditions (Table 6) using end-point titration after 20 hours of cellulolysis. The most sensitive and effective reagent was p-HBH, though even this reagent required enzyme denaturing reaction conditions, precluding its use in continuous assay of cellulase activity. As described in chapter 1, CelF is a GH family 48 cellobiohydrolase II that processively removes reducing sugar equivalents (cellobiose) from single chains of polymeric cellulose. However, due the intractable nature of crystalline cellulose, the rate of hydrolysis is very slow, on the order of 50 $mg \bullet L^{-1}hr^{-1}$ using micromolar enzyme concentrations. This slow rate of cellulolysis is problematic when using crystalliine cellulose as a model substrate. Demain and co-workers[8] have commented on the use of trinitrophenyl-functionalized carboxymethyl cellulose substrates as a more convenient substrate for select endoglucanases of *C. thermocellum*, although the broader application of this substrate to cellobiohydrolases has not, to our knowledge, been reported.

3.1.1.2 HPLC-based methods

High-performance liquid chromatography (HPLC) has been used to monitor the activity of cellulases against non-crystalline cellulose substrates.[192-194] In a discontinuous assay, aliquots of the soluble fraction of cellulolysis are analyzed by HPLC to quantify soluble cellodextrins and glucose. The method requires the use of specialized stationary

phases capable of resolving mixture of underivatized cellodextrins. At an analytical scale, typical resins include beta-cyclodextran, Pb^{2+} or Ca^{2+}-bound cation exchange resins, and high-resolution gel filtration media. Detection methods include electrospray mass spectrometry, refractive index detection (RID), and pulsed amperometric detection (PAD). Although LCMS is able to identify different sugar products, quantitation is at best problematic.[195] RID is the most common quantitative detection for the HPLC analysis of sugars. Amperometric detection, which measures an electrical response to reducing sugar oxidation over a gold or enzyme-linked anode, is the most quantitative detector of soluble sugar, but is not often used for HPLC detection.[194]

The HPLC-based method provides valuable information regarding cellulolysis intermediates and, thus, gives perhaps the best picture of the time course of cellulolysis, especially intermediate product distributions. On the other hand, HPLC is not ideal for routine functional analysis of cellulase activity due to slow hydrolysis rates and special instrumentation requirements. Sample collection is usually performed over 24 or more hours, and individual samples from the soluble fraction of a cellulolysis mixture require centrifugation prior to analysis. Samples are then individually analyzed by HPLC, a process that can take 30-60 minutes per sample. Individual peak integrations are finally fit to calibration curves for quantitation.

3.1.2 Real-time cellulase quantitation

3.1.2.1 Microgravimetric analysis of cellulases

Turon and co-workers have introduced an intriguing method of quantifying cellulose binding and hydrolysis in real time using a Quartz Crystal Microblance (QCM). The change in the frequency of an oscillating surface coated with cellulose was monitored over the time-course of cellulolysis and fit to an empirical model.[156] As the mass of the cellulose decreased over time the change in frequency (Δf) as a function of time (t) increased, relating frequency changes directly to enzyme activity via Equation 36.

$$\Delta f = \left\{ \begin{array}{c} M_{\max}(1 - e^{-t/\tau}), t < I \\ A + \dfrac{B - A}{\left(1 + e^{\left\{(V_{50} - t)/C\right\}}\right)}, t > I \end{array} \right\}$$

Equation 36

The t < I domain describes the early-stage curve fit as an exponential decay where frequency change is primarily indicative of enzyme binding to cellulose. M_{max} is the maximal observed frequency change in Hz, and τ is the reciprocal of the enzyme binding rate measured in minutes. During the t > I domain, Δf becomes a function of the enzymatic hydrolysis of cellulose, which Turon and co-workers fit a Boltzmann-sigmoidal expression containing fitting values A, B, and C. A corresponds to the minimum frequency in the sigmoidal curve, which is in turn related to the time period

between binding and hydrolysis. B is the maximal Δf value reached at the end of the hydrolysis reaction. V_{50} is the inflection point in the plot of Δf vs. t, describing the time at which maximal product conversion rate is reached. C is considered as the inverse of the 1st order hydrolysis rate constant.

Turon et al. fit the raw QCM data of endoglucanase-catalyzed cellulose hydrolysis to Equation 36. Although the fit was good, a biophysically meaningful interpretation of the fit parameters in terms of turnover number and a Michaelis constant was, at best, only qualitatively covered in the study; this was in part due to an inability to parse the contributions of surface hydration versus surface depolymerization in the raw data. The potential utility of microgravimetry for real-time cellulase quantitation is appealing and warrants further development.

3.1.2.2 AFM-imaging of cellulase action

Direct imaging of cellulase activity in real time has been accomplished using high-speed atomic force microscopy (HS-AFM).[196] In 2009, Igarashi et al. reported the processive movement of *T. reesei* CBH I Cel7A on crystalline cellulose fibers, and related the movement to the direct measurement of cellobiohydrolase activity.[197] They measured the continuous movement of CBH I across a cellulose fiber over 30-60 seconds. Enzyme mutation experiments proved the cellulase motion was indeed coupled to the celliobohydrolase activity. Removal of the non-catalytic CBM domain resulted in a decrease in cellulase-cellulose binding, though the rate of motion of the catalytic domain

(CD) was similar to the wild-type enzyme. The cellulase movement ceased upon mutation of the active site glutamate residue to catalytically inactive glutamine or mutation of the crucial CD sugar-binding residue W40 to alanine. The wild-type cellulase moved at an average speed of 4 ± 1 nm•s^{-1}, correlating to cellulose hydrolysis rate constant of 4 ± 1 s^{-1}, assuming a cellobiose length of about 1 nm.[197, 198]

The imaging approach provides significant chemical and kinetic insight into the direct cellulase action on insoluble cellulosic substrates, but the approach is limited to processive cellulases with well-defined substrate sites. As such, the use of HS-AFM for the study of endoglucanases, which comprise the majority of the cellulosomal cellulases, is far less valuable than its corresponding use to monitor cellobiohydrolase kinetics.

3.1.2.3 Calorimetric analysis of a cellulase-coupled enzyme hydrolysis

Calorimetric analysis of enzyme-catalyzed reaction rates is a powerful tool in quantitative biochemistry.[199] As described above (see 2.4.1 CMB3a-cellodextrin binding probed with Isothermal Titration Calorimetry), the ITC instrument applies power (dQ/dt) to maintain an isothermal relationship between the sample cell and a reference cell. If the sample cell contains a chemical reaction with a known reaction enthalpy, the power may be taken as a direct measure of the rate of the reaction:

$$v = \frac{d[P]}{dt} = \frac{dQ}{dt} \bullet \frac{1}{V_0 \Delta H_{app}}$$

Equation 37

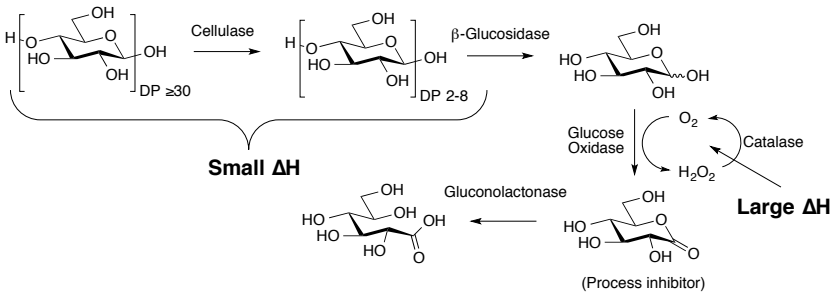
where v is the reaction rate, P is the reaction product, V_0 is the cell volume, and ΔH_{app} is the measured (or known) molar reaction enthalpy in kcal•mol^{-1}.

The direct calorimetric evaluation of cellulase kinetic parameters is not possible due to the inherent limitations of the instrument. Williams and Toone[199] detailed that the lower limit of k_{cat} for an enzyme-catalyzed reaction with a given ΔH_{app} and K_m is set by the instrument sensitivity such that

$$\left(\frac{k_{cat}}{K_m[E][S]}\right)\Delta H_{app}V_0 > 1\ \mu cal \bullet s^{-1}$$

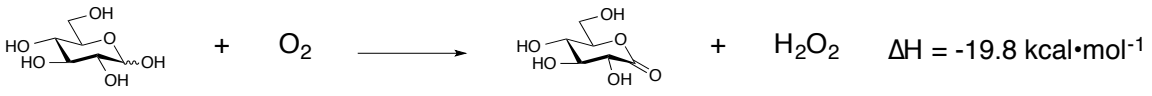
Equation 38.

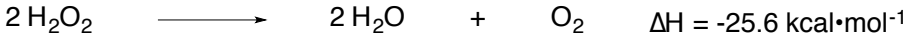
Cellulase-catalyzed hydrolysis of crystalline cellulose is very slow (k_{cat} ~0.2 – 40 min^{-1} for CBH II),[167] and the glycosidic bond hydrolysis enthalpies are small (~ -0.6 kcal•mol^{-1}). In order to overcome the lower limit of instrument sensitivity, Murphy et al. reported an enzyme-coupled system designed to enthalpically amplify the cellobiohydrolase activity in ITC.[167] As shown in Scheme 5, the principle product of cellobiohdroylase activity, cellobiose, was converted to D-glucose by almond β-glucosidase (BG). Glucose was then oxidized using the glucose oxidase (GOX) and catalase system, generating gluconolactone, which hydrolyzed spontaneously to gluconic acid. The GOX-catalyzed



Scheme 5: Cellulase signal amplification in ITC

oxidation of glucose requires O_2 as the terminal electron acceptor, yielding the byproduct H_2O_2, which is subsequently reduced to water by catalase to regenerate oxygen. Given the large reaction enthalpies of glucose oxidation and peroxide reduction, the enthalpy of cellobiohydrolase activity was magnified to a combined ΔH_{app} = -84.8 kcal•mol^{-1}. However, the rate of reaction was indicative of only cellobiohydrolase activity as cellulolysis was the rate-limiting step.





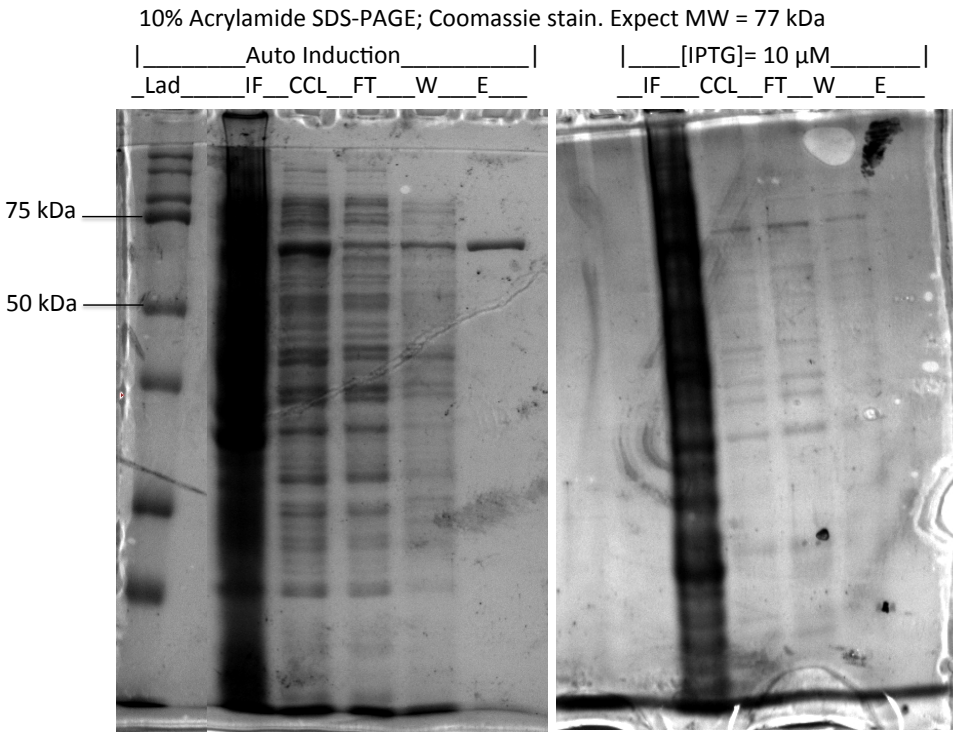
The direct oxidation of glucopyranose initially yields gluconolactone, which is a potent inhibitor of BG. Murphy et al. reported that BG-inhibition did not appear to slow the apparent reaction rate; however, our own initial studies indicated significant BG-inhibition.[167] Yeast strain gluconolactonase enzymes could be added to the reaction mixture to overcome the gluconolactone inhibition of BG.[200-202]

Calorimetric evaluation of cellulase activity provides the opportunity to monitor cellulose hydrolysis in a real-time, continuous fashion against essentially any lignocellulosic substrate. Due to the required enzyme-coupled enhancement of cellulolysis enthalpy, the observed reaction rates are indicative of cellobiose released and are only indirect measurements of cellulose depolymerization by endoglucanases. The direct measure of cellulose depolymerization has not, to our knowledge, been realized with ITC. It is likely, however, that endoglucanase kinetic profiles may be deduced with proper mathematical modeling of endoglucanase-induced changes in the apparent cellobiohydrolase activities.

3.2 Heterologous expression of chimeric Clostridial cellulases in E. coli

The GH family 9 endoglucanase CelG and GH family 48 cellobiohydrolase II CelF from *Clostridium cellulolyticum* were expressed in BL21(DE3) *E. coli* under the T7 RNA polymerase/ *lac* operon system as described in chapter 2. Expression plasmids pET9Gc, pET9Gt, pET48Fc, and pET48Ft were a generous gift from Ed Bayer at the Wizemann Institue in Rehovot, Israel and Henri-Pierre Fierobe from the University of Marsaille, France. The plasmids encode for two His_6-fused wild type enzymes (denoted Gc and Fc) that contain the *C. cellulolyticum* dockerin type I domain and two chimeric enzymes (denoted Gt and Ft) that contain an orthogonal C. thermocellum dockerin type I domain and C-terminal His_6 affinity tags (see section 1.1.3.2 Designer miniature cellulosomes).

Proteins were expressed and purified following modified literature procedures as detailed in chapter 5.[121, 122, 203-205] Our observations corroborated previous reports that CBH II CelF was efficiently expressed by induction with IPTG to a final concentration of 40 μM at 17 °C for 17 hours. CelG, howevder, did not express following induction with 10 μM IPTG as previously described.[122, 203] We explored the use of constitutive induction with glucose-supplemented lactose media (the so-called "auto-induction" protocol[206]) at 25 °C for 24 hours and observed a substantial improvement in CelG expression as determined by SDS-PAGE analysis of IMAC-purified cell lysates (Figure 38).



*IF = insoluble cell fraction; CCL = cleared crude lysate; FT = IMAC flow through; W = column wash fraction; E = IMAC elution with 500 mM imidazole.

Figure 38: SDS-PAGE analysis of Cel9Gc expression by auto-induction or with IPTG

Cel48Ft purified by Ni(II)-IMAC provided two elution bands, the latter of which was presumed to be a degradation product; autolytic activity of cellulosomal cellulases is commonly observed during expression and purification. The parent protein fraction was purified to homogeneity via size exclusion chromatography over an S100-HR column (GE Healthcare), as depicted in Figure 39.

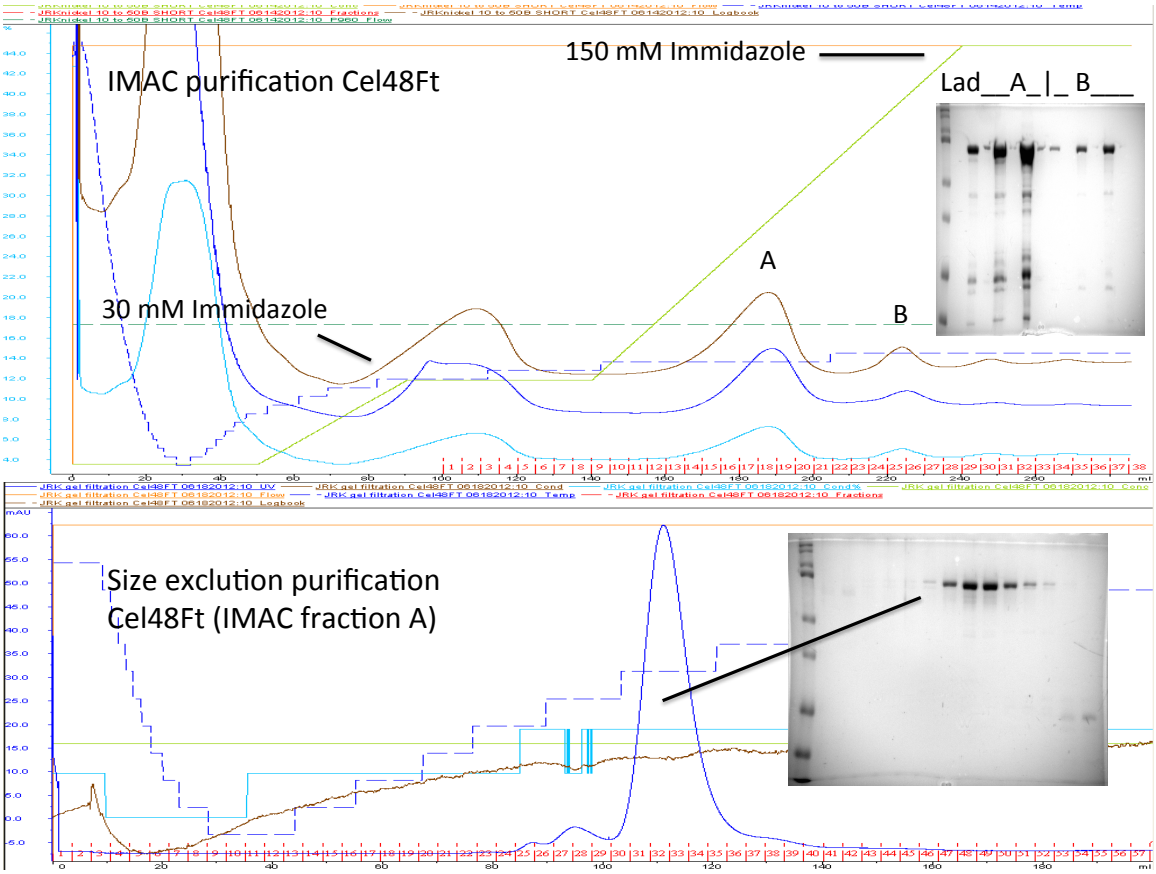


Figure 39: FPLC and SDS-PAGE traces of Cel48Ft purification

Cel9Gc was isolated from the soluble cell lysate fraction by direct elution over a Ni(II)-IMAC column. The partially purified fraction was further purified via anion exchange chromatography over the strong anion exchanger Q-sepharose (Figure 40). A minor

degradation product was apparent in the elution fractions. Time-dependent evalution of the eluent over several days indicated significant accumulation of this byproduct, presumably the result of autolysis. Gratifyingly, this autolysis was greatly slowed for both types of enzymes via the addition of the aspartyl protease inhibitor Pepstatin A at a concentration of 1 μg/ mL.

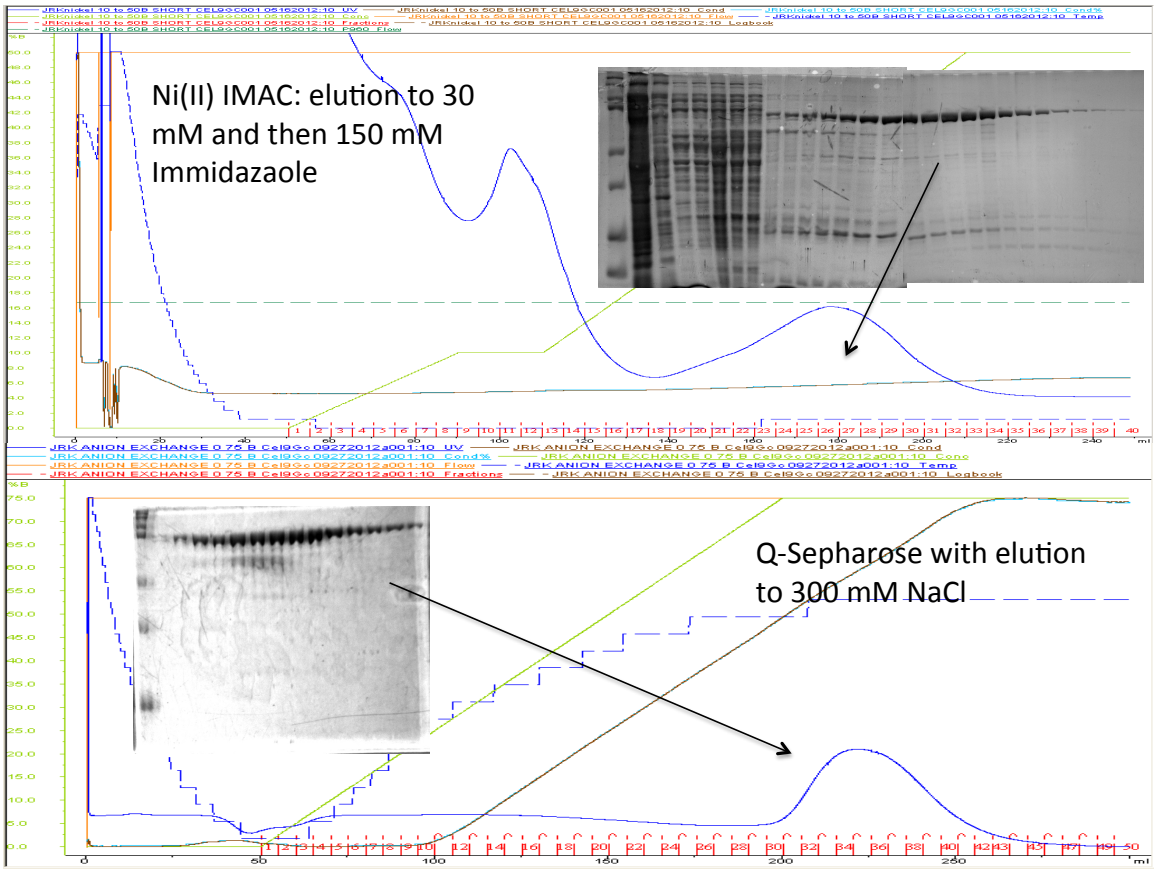


Figure 40: FPLC and SDS-PAGE traces of Cel9Gc purification

3.3 Towards the chemical synthesis of a fluorogenic cellohexaoside

3.3.1 Design and retrosynthesis of a flourogenic cellohexaoside

Glycan hydrolases are commonly assayed continually using *p*-nitrophenyl glycosides or similar chromophoric glycosides. Upon aglycone hydrolysis, *p*-nitrophenoxide (*p*NP) is released and enzyme activity is related to increasing absorbance over time. Cellulases, however, do not readily hydrolyze the aglycone unit from *p*NP-oligosaccharides; rather, the internal glycosidic bonds are cleaved, and no detectable *p*NP is realeased. The potential of fluorescence quenching to measure cellulase activity on native, polymeric cellulosic substrates in vitro and in vivo has been realized by Driquez et al.,[207, 208] who made cellodextrin-based fluorogenic probes of cellulase activity that are compatible with both the major classes of exocellulases (cellobiohydrolase I and II) and endocellulases. As described by Driguez et al., the cellohexaose core of **1** (Figure 41) increases substrate availability to all major types of cellulases, including processive endocellulases such as *Clostridium cellulolyticum* Cel9G. By incorporating the fluorophore 5-[(2-aminoethyl)amino]naphthalene-1-sulfonic acid (EDANS) and florescence quencher 4-[4-(dimethylamino)phenylazo]benzoic acid (DABCYL) in close proximity on the ends of the cellohexaose core, Förster resonance energy transfer (FRET) is observed, causing minimal detectable florescence emission from EDANS. Upon cellulase-catalyzed hydrolysis of the saccharide core, the FRET pairs separate, and EDANS emission is related to enzyme activity in real time.

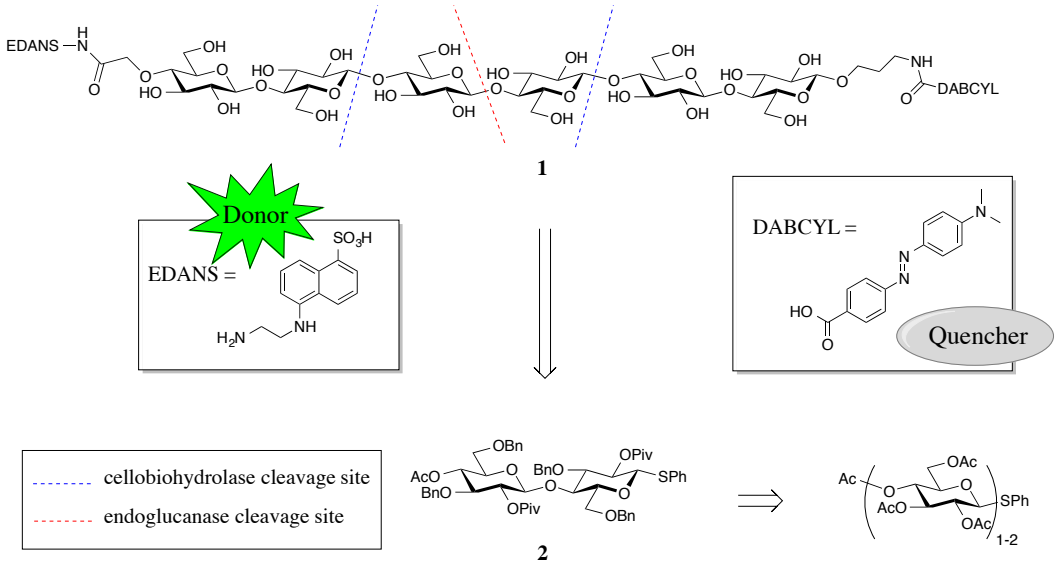
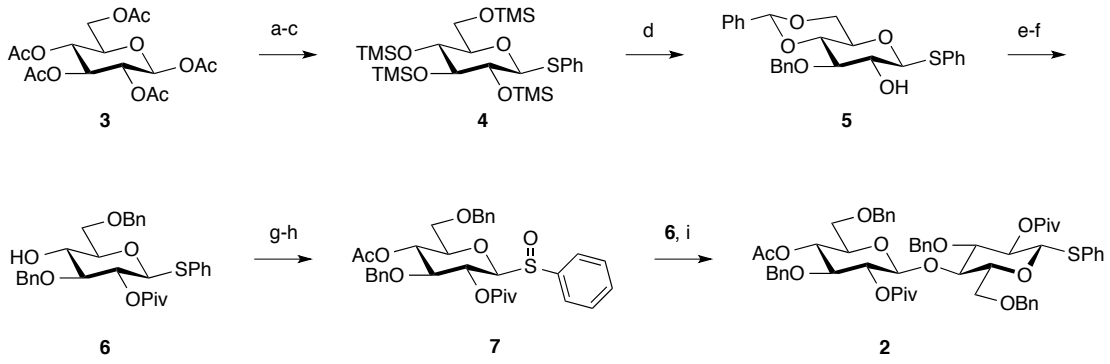


Figure 41: Design of fluorogenic cellohexaoside 1

Although substrates like **1** are effective as real-time, sensitive, universal probes of cellulose activity, the field has been reluctant to adopt these substrates in cellulase assays, presumably due to the poor availability of these probes. Previous syntheses of fluorogenic probes detailed a chemoenzymatic route, highlighted by the coupling of glucosyl fluoride donors to underivatized cellobioside acceptors using mutant cellulase catalyst Cel7a E197A from *Humicola insolens*.[207, 208] To provide a more accessible route to **1**, we developed a *de novo* chemical synthesis via the key disaccharide building block **2**. This key intermediate was easily prepared on a multi-gram scale from commercial starting materials in either eight steps from glucose pentaacetate or in three steps from thiophenyl heptaacetyl-β-D-cellobioside. Disaccharide **2** was efficiently oligomerized in a step-wise fashion to form the differentially protected hexasaccharide with orthogonally reactive handles at opposite termini. Finally, the late-stage, site-specific

incorporation of the FRET-paired dyes at these handles allows for the incorporation of fluorophore and quencher molecules.

3.3.2 Two-pronged approach to synthesize crucial disaccharide building block 2



a) PhSH, BF_3•Et_2O, CH_2Cl_2, 0°C-->rt, 88% yield; b) cat. $NaOCH_3$, CH_3OH; c) TMSCl, pyridine, 98% yield over two steps; d) $Cu(OTf)_2$,CH_2Cl_2, CH_3CN, PhCHO, 0°C-->rt, then Et_3SiH, 0°C, then TBAF, THF, $NaHCO_3$, rt, 85% yield; e) PivCl, DMAP, Et_3N, CH_2Cl_2, 0°C-->rt, 97% yield; f) $NaCNBH_3$, HCl,4ÅMS, dioxane, THF, 0°C-->rt, 92% yield; g) Ac_2O, DMAP, Et_3N, CH_2Cl_2, rt, 100% yield; h) mCPBA, CH_2Cl_2, -10°C, 98% yield; i) DTBMP, Tf_2O, 4ÅMS, CH_2Cl_2, -78°C--> -60°C, then **5**, -60°C--> -35°C, 61% yield.

Scheme 6: Synthesis of disaccharide building block 2

Commercially-available β-D-glucose pentaacetate (**3**) was converted to the thiophenyl glycoside in the presence of the Lewis acid catalyst BF3•Et2O and recrystallized as the β diasteriomer in 88% yield.* The acetate functional groups were removed via Zemplen de-acetylation and the crystalline product was silylated with TMSCl in pyridine to yield persilylated thiophenyl glycoside **4** in a 98% yield (two steps). In a key single-pot transformation, **4** was elaborated into the regioselectively

* Interestingly, no reaction was observed with the α-acetate was used (results not shown).

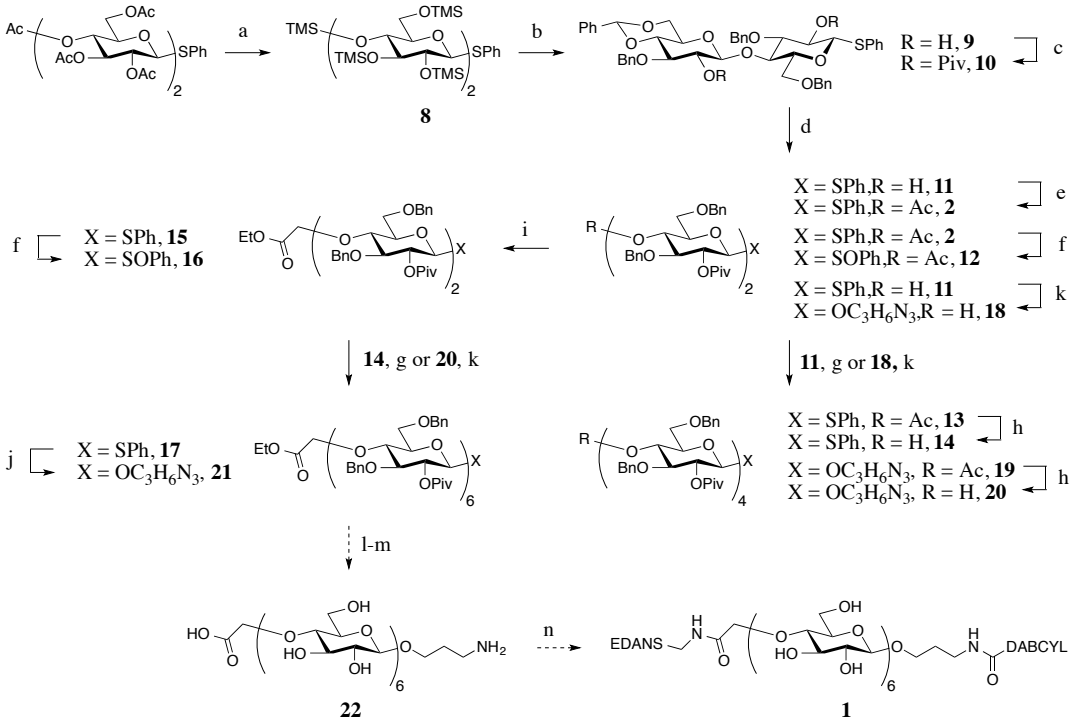
protected 4,6-0-benzylidene-3-O-benzyl-2-hydroxy-β-thiophenylglycoside **5** in 85% yield (72% crystalline yield) using a slight modification to the copper(II) trifluormethanesulfonate method of Beau and co-workers.[209] **5** was pivaloylated with trimethylacetyl chloride in the presence of stoichiometric DMAP, and a subsequent regioselective reductive opening of the 4,6-O-benzylidene acetal with $NaCNBH_3$ and dry HCl yielded C4-hydroxy nucleophile **6** in 92% yield. Thus, the C4-OH nucleophile required to create the (1→4) glycosidic bond and the C2-O-pivaloyl participatory group that generates the 1,2-trans geometry (β) of the glycosidic linkage were instilled in a regioselective manner in high yield over just three steps.

Given the unsatisfactory glycosylation yields obtained with unreactive glycosyl fluoride and glycosyl trichloroacetimidate (35% yield) donors, we investigated the sulfoxide activation strategy of Kahne and co-workers.[210] **6** was converted to the sulfoxide glycosyl donor **7** upon quantitative acetylation and subsequent oxidation with mCPBA in sufficient purity for glycosylation reactions after a simple aqueous workup. The chemoselective activation of the glycosyl sulfoxide **7** in the presence of the thiophenyl glycoside acceptor **6** provided disaccharide **2** in 61% isolated yield.

Although disaccharide **2** was readily synthesized in a total yield of 39% over 9 steps encompassing only 2 column purifications, the total yield with respect to saccharide carbon input was a disappointing 9.7% yield. The loss in valuable saccharide carbon is primarily due to the un-recoverable donor, which was used in excess during

the glycoside coupling reaction. We thus explored an alternative synthesis of **2** beginning from cellobiose. Thiophenyl heptaacetyl-β-D-cellobioside, which is both commercially available (Abacipharm Corporation; Columbia, MD, USA) and readily synthesized via known procedures,[211] was chosen as an appropriate alternative starting point.

The shortened synthetic route to the intermediate disaccharide **11** and further elaboration into the desired hexasaccharide cellulose substrate is outlined in Scheme 7. Thiophenyl helptaacetyl-β-D-cellobioside was subjected to Zemplen de-acetylation, and the crude material was directly silylated with TMSCl to yield thiophenyl persilyl-cellobioside **8**. The copper-catalyzed reductive benzylation of **8** yielded C2,C2'-diol **9** in 40% isolated yield. Initial attempts to pivaloylate **9** using PivCl and DMAP with Et_3N were unsatisfactory, do to the accumulation of a poorly-reactive mono-pivaloyl intermediate. An alternate method[212] using catalytic scandium(III) trifluoromethanesulfonate and excess pivalic anhydride cleanly produced the desired di-pivaloyl product **10** in 80% yield in only 2 hours at 0°C. A subsequent reductive opening of the 4',6'-*O*-benzylidene acetal with sodium cyanoborohydride and dry HCl yielded disaccharide alcohol **11** in 90% yield.



a) NaOMe/ MeOH, then TMSCl, immidazole, Et_3N; b) $Cu(OTf)_2$ (0.06 eq), PhCHO (4.4 eq), DCM:ACN (4:1), 0 °C, then Et_3SiH (2.1 eq), 0 °C; then TBAF (4.4 eq), $NaHCO_3$ (4.4 eq), THF, rt; c) $Sc(OTf)_3$, Piv_2O, ACN; d) $NaCNBH_3$, HCl, Dioxane/ THF, 0 °C - rt; e) Ac_2O, DMAP, Et_3N, DCM; f) mCPBA, DCM, -15 °C; g) 4ÅMS, Tf_2O, DTBMP, DCM, -60 °C --> -20 °C; h) Guanidine, cat. NaOMe, EtOH/ DCM; i) NaH, DMF, -20 °C --> 0 °C, then $BrCH_2CO_2Et$; j) $HOCH_2CH_2CH_2N_3$, 4ÅMS, I_2, AgOTf, DCM, 0 °C --> rt; k) $HOCH_2CH_2CH_2N_3$, 4ÅMS, NIS, AgOTf, DCM, -20 °C --> 20 °C; l-n) deprotection and dye ligation.

Scheme 7: Synthesis of cellohexaoside 1

3.3.3 Synthesis of cellohexaoside 21

An orthogonal block strategy to obtain the larger tetra- and hexasaccharide intermediates from the common disaccharide building block **11** is outlined in Scheme 7. Acetylation of **11** as described for compound **6a** afforded compound **2** in quantitative yield. mCPBA oxidation of **2** afforded disaccharide donor **12** as a mixture of sulfoxide diastereomers. The donor mixture was activated with triflic anhydride and coupled to

glycosyl acceptor **11** as for **2** (Scheme 6) to yield tetrasaccharide **13** in 60—65% yield with ~20-30% yield of the unwanted aglycone transfer side-reaction to regenerate compound **2**. A chemoselective solvolysis of the C4′ acetoxy group of **13** using catalytic methoxide in a guanidine-buffered ethanol/dichloromethane mixture (7:1) yielded tetrasaccharide acceptor **14** in nearly quantitative yield.

A carboxymethyl handle for late-stage incorporation of the EDANS dye onto the hexasaccharide core was introduced by way of the disaccharide donor **16**. **16** was synthesized via an initial *O*-alkylation of alcohol **11** with sodium hydride and ethyl bromoacetate to form compound **15** and subsequent oxidation of **15** to yield **16** as a mixture of sulfoxide diastereomers in 80% yield over two steps. The hexasaccharide core was synthesized via triflic anhydride activation of **16** and coupled to glycosyl acceptor **14** to yield hexassaccharide **17** in moderate yield (45—55%).

The absolute configurations of the β-1,4 glycosidic linkages in oligosaccharides **2**, **13**, and **17** were confirmed by high resolution ^{1}H-NMR using 3-bond couplings of the anomeric protons ($J_{H1\text{-}H2}$ = 8-11 Hz, 800 MHz; experimental chapter 5). Anomeric protons were assigned using 2D correlation experiments: (^{1}H/^{1}H)COSY for disaccharide **2** and (^{1}H/^{13}C)HMQC for oligosaccharides **13** and **17** (see chapter 5 and Appendix B).

Numerous attempts to elaborate thiophenyl cellohexaoside **17** into the azido glycosyl ether **21** using NIS/ AgOTf, NIS/ TMSOTf, NIS/ BF_3•Et_2O, or IDCP failed to produce any reaction. The use of activated iodine reagents such as ICl or I_2/ AgOTf gave

better results, although isolated yields remained low (**21**, 30%). Due to the poor reactivity of the thiophenyl glycoside with iodonium reagents the BSP/ Tf_2O and Tf_2O/ glycosyl sulfoxide methods were explored. Though these promoters efficiently activated the thiophenyl aglycone, the intermediate glycosyl triflates were likely unstable at the higher temperatures (> -30°C) needed to dissolve the hexasaccharide, and the formation of a highly stable ortho ester product precluded the formation of the desired azido glycosyl ether, even after addition of stoichiometric TMSOTf.

Attempts to elaborate disaccharide **11** into the azido glycosyl ether **18** proved effective with NIS/ AgOTf at room temperature (70-90% yields); thus, the disaccharide building block **18** was used to produce the desired hexasaccharide intermediate **21. 18** was coupled to thiophenyl glycoside donor **2** with NIS and AgOTf to form tetrasaccharide intermediate **19**, which was de-acetylated with guanidine-buffered ethoxide to yield tetrasaccharide alcohol **20**. **20** was coupled to thiophenyl glycoside donor **2** using the NIS/AgOTf method as described above for the formation of compounds **18** and **19** (Scheme 7).

3.3.4 Future efforts towards the synthesis of cellohexaoside 1

Protected hexasaccharide **21** will be globally de-protected using standard techniques to afford compound **22** over a two step sequence. An NHS-activated DABCYL molecule will be ligated to the amino terminus of **22** and, subsequently, an amine-terminated EDANS dye will be incorporated at the carboxy terminus of **22** using

EDC/coupling conditions in aqueous milieu to generate the desired cellulase substrate **1** in a single pot. The full conversion of commercially available thiophenyl heptaacetyl-β-D-cellobioside to **1** is projected to be completed in 13 steps total.

3.4 Conclusions

The kinetic characterization of cellulases requires robust and reliable assays to quantify functional cellulase content of from recombinant enzyme preparations. The state-of-the art for discontinuous analysis of cellulase activity include reducing sugar chromogenic assays and HPLC methods. These methods are time consuming and not capable of assaying cellulase function in real time. Continuous cellulase assays include microgravimetry, high-speed AFM, and microcalorimetry. Though these methods are capable of real time and, at times, direct observation of cellulase activity, they are not suitabley high-throughput for routine cellulase assays.

The design and synthesis of a cellohexaoside FRET-based probe for real-time quantitation of cellulase activity with flouresence spectrometry was proposed. The flouorogenic cellohexaoside probe is nearly identical to one was previously reported by Driguez and co-workers.[208] The chemo-enzymatic synthesis presented by Driguez en route to the water-soluble flourogenic cellohexaoside included 14 steps from a known synthetic intermediate (hexaacetyl lactose). The chemo-enzymatic route highlighted the use of a non-commercial, mutant cellulase catalyst to create the key glycosidic linkages in the reaction of glycosyl fluoride donors with un-protected acceptors. Unfortunately,

no reports have surfaced from other laboratories that have used this synthesis to prepare cellulase activity probes. We hypothesized that an equally concise, but fully chemical synthesis of the cellohexaoside might provide a more scalable and general method to enable the wide-spread production of the fluorogenic cellohexaoside.

Here, a fully chemical synthesis of cellohexoaside **1** in 13 steps was presented. Two routes to synthesize a key thiophenyl glycoside building block **2** were presented, with the more concise route providing the disaccharide in 4 steps from a commercial starting material. The disaccharide was coupled in an iterative fashion to glycosyl sulfoxide donors under the chemoselective activation conditions of Kahne and co-workers. However, the poor reactivity of the thiophenyl cellohexaoside **17** precluded the elaboration of **17** with dye-reactive functional groups on the hexasaccharide core. Thus, an alternate route was created in which the dye-reactive handles were originally incorporated into the disaccharide building blocks prior to the production of the cellohexaoside. The disaccharide blocks were activated with electrophilic iodonium (NIS/AgOTf) to yield the cellohexaoside **21**. Elaboration of cellohexaoside **21** into the underivatized FRET-paired hexasaccharide **1** and its subsequent use in cellulase activity assays is a continued goal of this work.

4. Enthalpic signature of methonium desolvation revealed using a synthetic host-guest system based on cucurbit[7]uril.[213, 214]†

Trimethylammonium (R-N^+Me_3; methonium) is an amphiphilic cation broadly distributed in biology, playing roles in processes as diverse as neurotransmission and lipid bilayer formation. Methonium-binding proteins often include binding sites that segregate the positively charged quaternary ammonium ion from bulk water.[215-219] Although many metal-binding proteins sequester inorganic cations using electron-rich residues such as histidine or ionized organic acids such as aspartate and glutamate,[220, 221] binding sites that accommodate methonium typically lack anionic moieties capable of forming strong electrostatic interactions.[216-219, 222] Rather, the most prevalent structural motif observed in methonium binding sites of proteins is an aromatic cage that forms cation-π interactions with the ligand.[223]

The means by which proteins bind methonium with a net favorable free energy is unclear. Cation-π interactions are weak: calculated and experimental values for the gas-phase interaction energy between tetramethylammonium (TMA^+) – a surrogate for methonium in many biophysical studies – and benzene range from –4 to –9 kcal•mol^{-1}.[224-227] In contrast to these modest values, the measured desolvation free energy of TMA^+

† Work presented in this chapter was performed in collaboration with Drs. Yi Wang, Pan Wu, and David Beratan and reprinted with permission from {213. Wang, Y., King, J. R., Wu, P., Pelzman, D. L., Beratan, D. N., and Toone, E. J. (2013) Enthalpic signature of methonium desolvation revealed in a synthetic host-guest system based on cucurbit[7]uril, *J Am Chem Soc 135*, 6084-6091.}. Copyright {2013}.

is +38.3 kcal•mol^{-1}, dominated by a large unfavorable desolvation enthalpy of +49.3 kcal•mol^{-1} at 298 K[228, 229] suggesting that the transfer of methonium from water to a hydrophobic binding pocket should produce large unfavorable enthalpic and free energetic contributions to binding. Even with four aromatic residues, the maximum number of aromatic rings geometrically capable of forming direct contact with methonium,[230] the energy available from cation-π interactions is too small to offset the +49.3 kcal•mol^{-1} enthalpic cost of methonium desolvation.

In contrast to gas phase-aqueous transfer thermodynamic parameters for TMA^+, a number of recent biophysical studies suggest that both free TMA^+ and methonium may be only weakly solvated.[231, 232] Hulme and coworkers studied the hydration of acetylcholine and observed a water structure that precludes significant charge-dipole interaction between water and the methonium group.[233] Rather, charge transfer to the methyl hydrogen atoms produces a large, diffuse charged species that interacts only weakly with water.[234, 235] Further, sites designed to accommodate methonium are unable to accommodate similarly sized inorganic cations. Thus, for example, acetylcholine esterase crystals soaked in CsCl show no evidence of Cs^+ occupancy in the binding pocket,[236] despite the fact that Cs^+ and TMA^+ have similar sizes[237] and similar gas phase-water transfer thermodynamics.[228, 229] Rather, the authors suggested that weak methonium solvation relative to inorganic cations may facilitate desolvation and binding of methonium in a hydrophobic binding site.

Despite compelling structural evidence from neutron scattering studies that methonium may in fact be only loosely bound to water,[233] thermodynamic evidence for such weak interactions is absent. Simple binding studies of TMA^+ and acetylcholine binding to proteins or synthetic hosts yield only aggregate binding thermodynamic data that contain contributions from both desolvation and solute-solute interactions of the entire ligand with the host. These data are of limited use for quantifying the thermodynamic consequences of desolvating individual functional groups, as pointed out by Diederich and others.[235, 238-240]

4.1 Isolating methonium desolvation thermodynamics by synthetic design

Two major tasks are required to quantify the thermodynamic consequence of methonium group desolvation upon encapsulation in a simple molecular cavity: 1) effectively separate the methonium-host binding thermodynamics from the contributions of the rest of the ligand; 2) determine the thermodynamics of solute-solute interactions separately from the desolvation effects.[241, 242] To achieve the first goal, we devised an experimental system comprising a synthetic host, namely cucurbit[7]uril (CB[7], Figure 42a), and a series of guests in which the structure of the bound complex and, consequently, the extent of ligand desolvation can be systematically altered by synthetic design. CB[7], a model receptor for protein-ligand complexes, sequesters methonium from aqueous solvent in a rigid, deep hydrophobic cavity.[235, 241, 243-245]

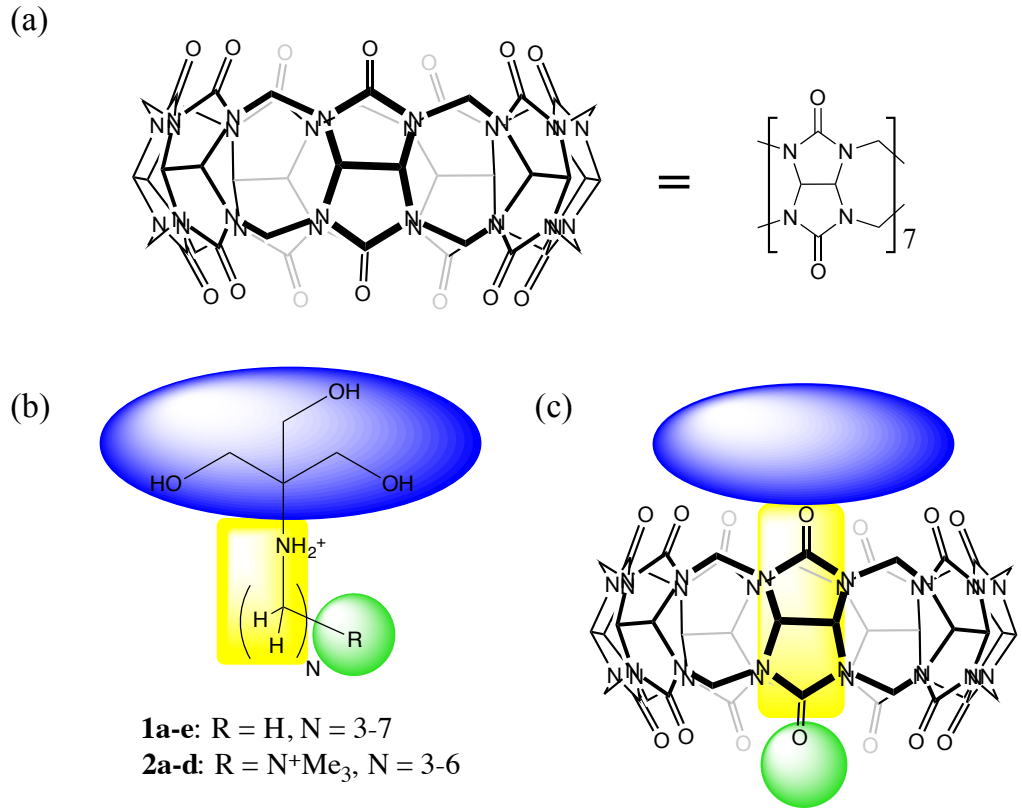


Figure 42: Molecular system to study methonium (green ball) desolvation. a) Cucurbit[7]uril structure; b) Ligand design containing anchor group (blue oval), variable length alkyl tether (yellow rectangle) and methonium; c) CB[7]: ligand complex

The value of CB[7] for probing desolvation has been detailed by Mock and others.[246-250] In our approach here, we use two ligand series (Figure 42b): (i) a reference series of 2-(hydroxymethyl)-2-(alkylamino)propane-1,3-diols (**1**,) and (ii) a test series 2-((1,3-dihydroxy-2-(hydroxymethyl)propan-2-yl)amino)-N,N,N-trimethylalkaminium bromides (**2**).[213, 214] Ligand series **2**, comprising an anchor (Tris), a variable length tether

(linear alkyl chain), and methonium, are bound to the host through the central cavity as depicted in Figure 42c. With a tether of sufficient length, steric effects preclude sequestration of the methonium group within the cavity, but rather leave the charged organic cation outside of the CB[7] cavity and at least partially solvated.

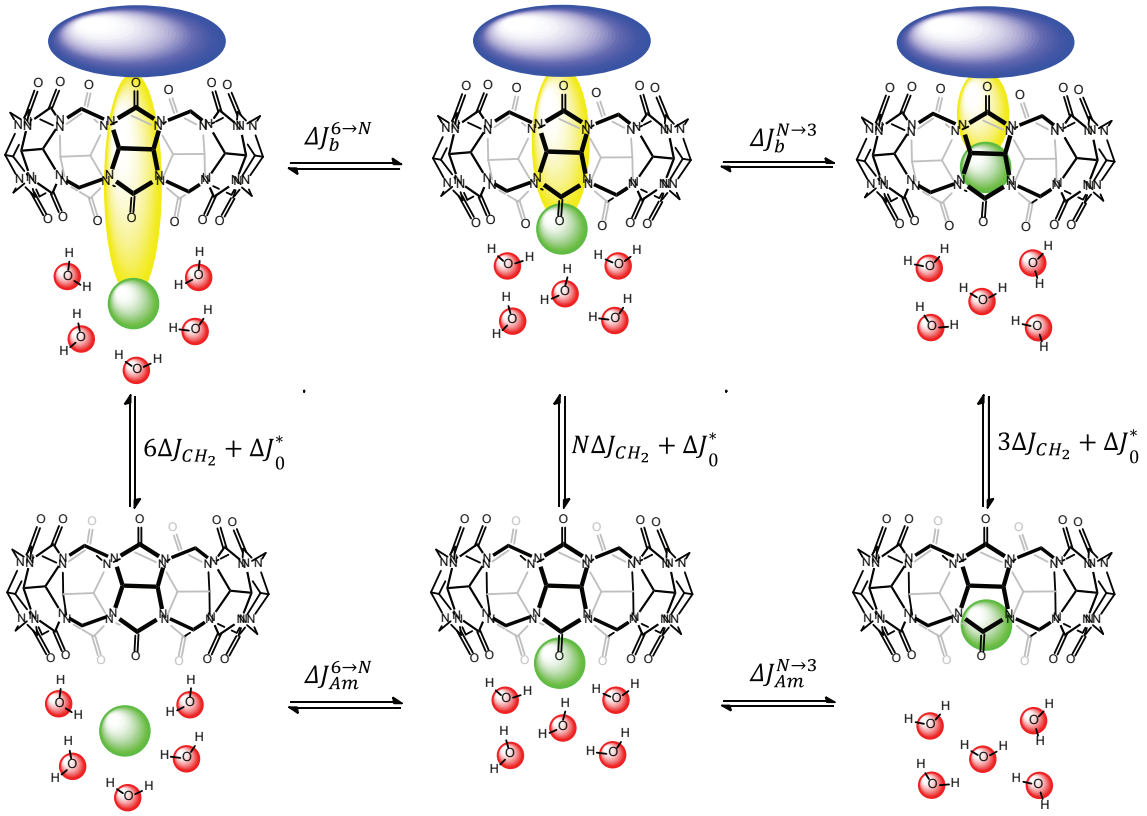


Figure 43: Thermodynamic cycle showing the isolation of methonium binding thermodynamics ($\Delta\Delta J_{Am}$)

Introducing the anchor group locks the entire ligand with respect to the host, and systematic shortening of the tether gradually repositions the methonium group from water to the host interior, desolvating the epitope (Figure 43).[246, 250-252] We conducted

thermodynamic analysis of CB[7]•**2** binding using isothermal titration microcalorimetry (ITC).

Following the additivity principle of Jencks and Page,[253, 254] binding thermodynamic parameters as a function of linker length yield a thermodynamic difference ($\Delta\Delta J$, where J = G, H, S, C_P) where $\Delta\Delta J_b$ is composed of two terms: a term attributable to linker variation ($\Delta\Delta J_{b,linker}$), and a term attributable to variation in the positioning of the terminal group, in this case methonium ($\Delta\Delta J_{b,Am}$). As the anchor moiety locks all ligands in a similar geometry when bound to the host, $\Delta\Delta J_{b,Am}$ can be extracted by comparison of ligand series **2** with a reference series of methyl-terminated ligands. Lacking the terminal methonium group, $\Delta\Delta J_b$ across the reference series essentially arises completely from $\Delta\Delta J_{b,linker}$. Thermodynamic analysis of ligand series **1** binding to CB[7] thus facilitates determination of the net energetic consequences of partitioning methonium between water and the non-polar cavity of CB[7] ($\Delta\Delta J_{b,Am}$), a feat unachievable with homobifunctional ligands such as the bisammonium alkanes used by Mock and Shih.[246]

Currently available approaches to the second challenge—separation of solute-solute from solute-solvent binding thermodynamics—typically involve *in silico* estimation of the solute-solute interaction free energies and enthalpies, discounting solute-solvent interactions.[255, 256] Thermodynamic analyses of various methonium-binding proteins suggest that binding free energies are dominated by the enthalpic

component.[215, 217, 223] Therefore, as an initial step towards quantifying methonium desolvation thermodynamic parameters, we used MD simulations to calculate the net change in methonium-CB[7] interaction enthalpy upon the incremental internalization of methonium ($\Delta\Delta H_{\mathrm{int},Am}$). Combining $\Delta\Delta H_{\mathrm{int},Am}$ with $\Delta\Delta H_{b,Am}$ obtained from ITC experiments, we estimate the enthalpic signature of methonium desolvation upon binding:

$$\Delta\Delta H_{desolv,Am} = \Delta\Delta H_{b,Am} - \Delta\Delta H_{\mathrm{int},Am}$$

Equation 39

We report the result of combined experimental and computational study here. Our results thow that the methonium group is only weakly solvated, incurring a smaller enthalpic penalty for desolvation than those derived from gas-phase studies of TMA^+, a finding consistent with the suggestion of Hulme and co-workers.[233] Our results also offer a cautionary prescription for using standard desolvation thermodynamic parameters based on free homologues to describe the behavior of tethered functional groups, and sheds light on the biophysical nature of choline-binding proteins and the energy landscape of epitope desolvation.

4.2 Thermodynamic model

We utilized a two-pronged approach to make an estimate of the desolvation enthalpy of methonium: 1) we separate the net thermodynamic effect of methonium encapsulation in the CB[7] cavity from the rest of the ligand matter; 2) we calculate the

change in methonium-CB[7] interaction enthalpy in order to back out the solvation terms from experimental ITC results. Our strategy for the first step follows from the anchor principle of group additivity,[257, 258] and the linear free energy relationship theory of Schneider for synthetic host-guest complexes,[259] which postulates that contributions to the binding thermodynamic parameters (ΔJ_b, J = G, H, S, Cp) from the anchor, the linker and the methonium group are additive. Our thermodynamic models are described briefly here; a more detailed derivation is provided in the Supporting Information. Our model is illustrated in Figure 43, which shows that the binding of ligand series **2** to CB[7] can be effectively transformed into the binding of methonium to a pre-desolvated CB[7] cavity.[260]

The CB[7]•**1** complexes were used to quantify the thermodynamic contribution of a methylene group, ΔJ_{CH_2}, to the overall binding thermodynamics. We applied a linear approximation by partitioning binding thermodynamic parameters as functions of the number of methylene groups in the linker (N):

$$\Delta J_b^{CB[7]\bullet 1} = N\,\Delta J_{CH_2} + \Delta J_0$$

Equation 40

where ΔJ_{CH_2} is the thermodynamic effect per methylene group in the alkyl linker and can be determined numerically by data fitting.

We next define the difference in binding thermodynamics ($\Delta\Delta J_b^{6\to N}$) between ligands **2a-c** with a linker of three to five methylene groups ($\Delta J_b^{CB[7]\bullet 2a-c}$) and the

reference ligand **2d** which contains a six methylene linker ($\Delta J_b^{CB[7]\bullet 2d}$), defining the CB[7]•**2d** complex as the reference state. Since the Tris anchor fixes the rest of the complex at a constant position, $\Delta\Delta J_b^{6\to N}$ contains contributions from the change in linker length, $(N-6)\Delta J_{CH_2}$, as well as a contribution associated with repositioning the methonium group from its equilibrium position in the CB[7]•**2d** complex to the corresponding position in the CB[7]•**2a-c** complexes ($\Delta\Delta J_{b,Am}^{6\to N}$, previously denoted as $\Delta\Delta J_{b,Am}$):

$$\Delta\Delta J_{b,Am}^{6\to N} = \Delta\Delta J_b^{6\to N} - (N-6)\Delta J_{CH_2}$$

Equation 41

$\Delta\Delta J_{b,Am}^{6\to N}$ consists of a desolvation term ($\Delta\Delta J_{desolv,Am}^{6\to N}$) and terms representing contributions from intrinsic interactions ($\Delta\Delta J_{\mathrm{int},Am}^{6\to N}$):

$$\Delta\Delta J_{b,Am}^{6\to N} = \Delta\Delta J_{\mathrm{int},Am}^{6\to N} + \Delta\Delta J_{desolv,Am}^{6\to N}$$

Equation 42

In this study, $\Delta\Delta H_{\mathrm{int},Am}^{6\to N}$ (previously denoted as $\Delta\Delta H_{\mathrm{int},Am}$) was calculated via MD simulation based on Equation 63 of Appendix C. Therefore, by experimentally determining $\Delta\Delta H_{\mathrm{int},Am}^{6\to N}$ *via* ITC, we can estimate $\Delta\Delta H_{desolv,Am}^{6\to N}$, that is the *net enthalpy change* for the incremental desolvation of methonium. Standard deviations of all terms were calculated by the propagation of experimental standard deviations.

4.3 Structural validation of the CB[7]•1 and CB[7]•2 systems

To implement our strategy of controllable internalization, we require an anchor moiety that cannot enter the central pore of CB[7], but that requires insertion of the linker and terminal epitope of the guest within the cavity. CB[7] perturbs the chemical shifts of bound protons, a well-established phenomenon, and provides a direct correlation between the ^{1}H-NMR complexation induced chemical shift ($\Delta\delta$) and the degree of internaliztion of these protons within CB[7].[243, 245, 261, 262] A combination of ^{1}H-NMR structural data and synthetic feasibility constraints supported the use of Tris as the anchor group.[214] The free and bound ^{1}H-NMR spectra of **1a-e** and **2a-d** are shown in Figure 44- Figure 52. The hydroxymethyl protons of ligands **2a-d** show a consistent down-field shift ($\Delta\delta$ = 0.21±0.03 ppm) upon binding, consistent with placement of the anchor group outside of the cavity in a constant position across the ligand series (Table 7 and Figure 54). The observed NMR shifts are in good agreement with those reported by Zhao and colleagues for alkylimidazolium ligands.[251] On the other hand, methonium protons undergo an increasing upfield shift as the alkyl linker is shortened, confirming that the locked position of the Tris anchor "pulls" the X-group into the cavity. Hydroxymethyl protons in the CB[7]•**1b-e** complexes consistently shift down-field upon binding in a fashion similar to those of the CB[7]•**2** complexes (Table 7 and Figure 54), highlighting again the powerful ability of the Tris anchor group to define the positions

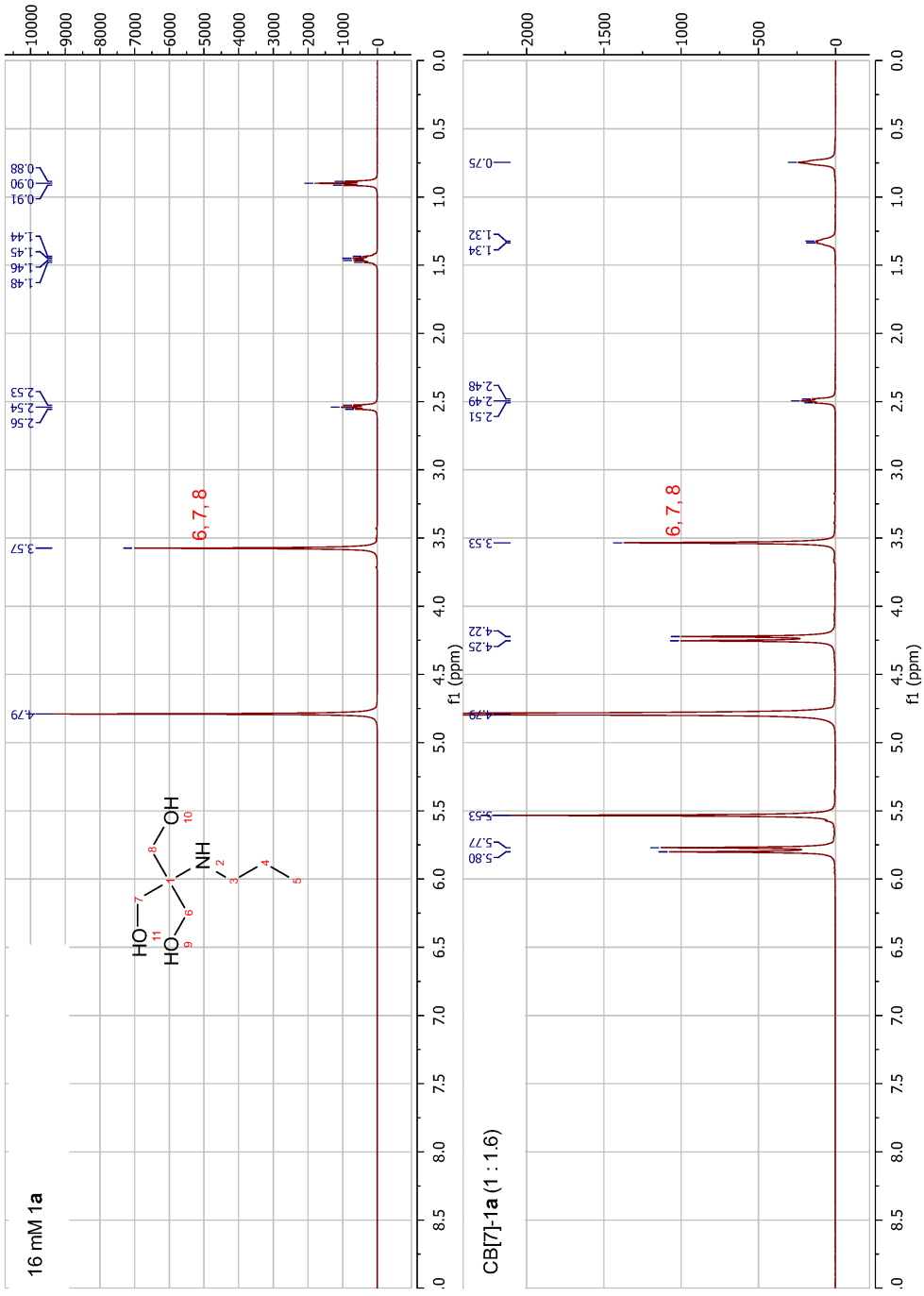


Figure 44: CB[7]•1a ^{1}H-NMR

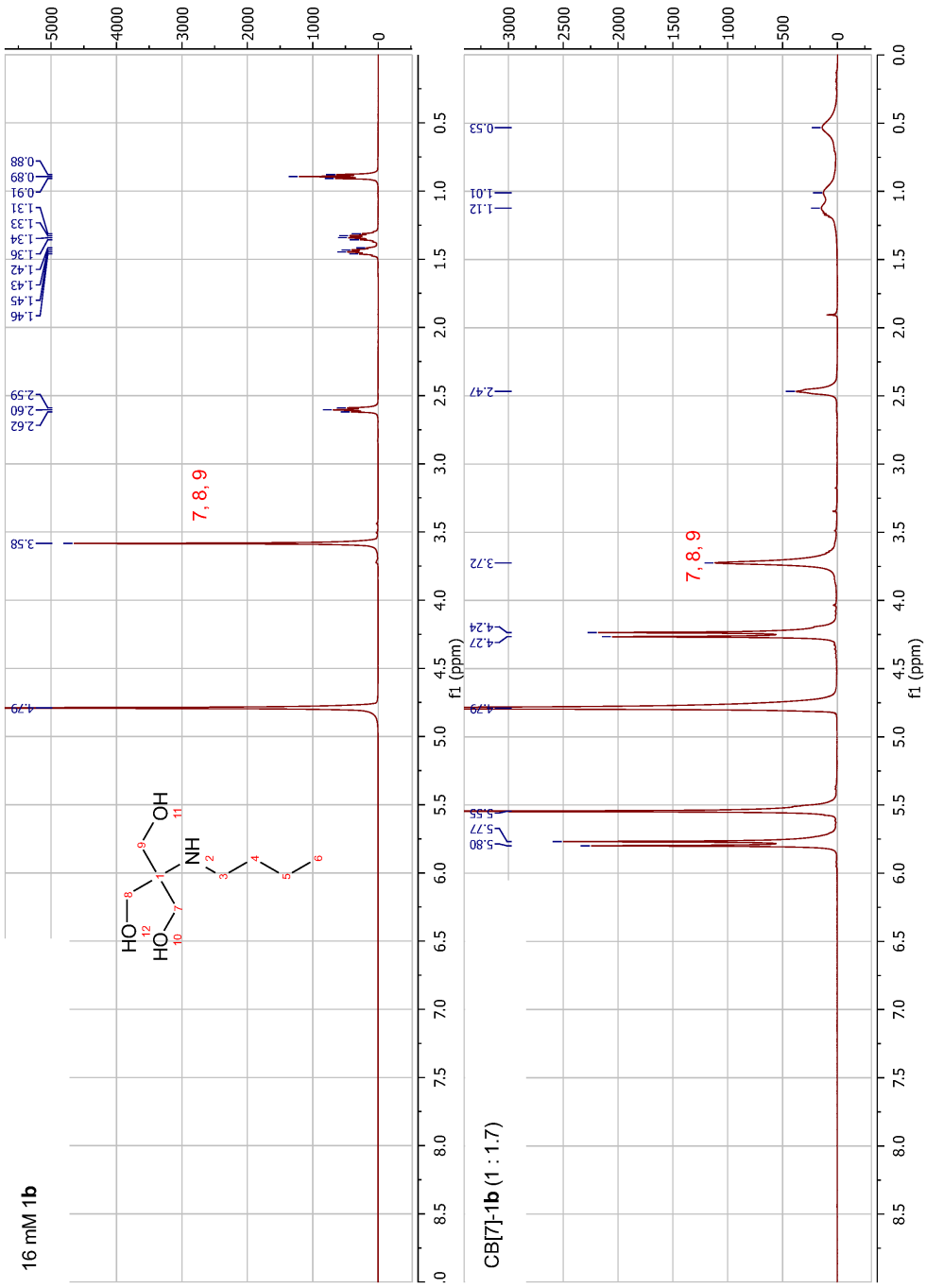


Figure 45: CB[7]•1b ^{1}H-NMR

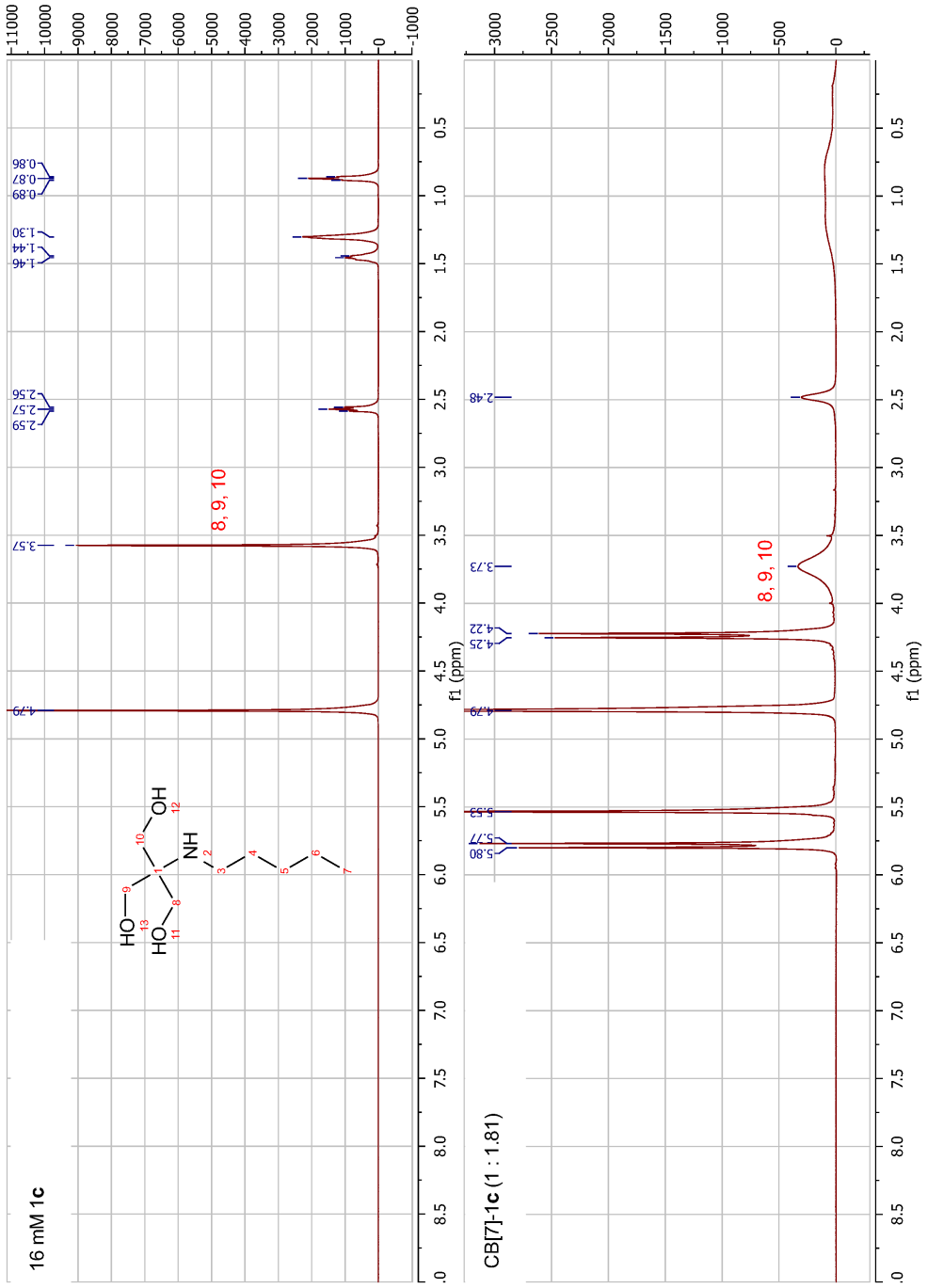


Figure 46: CB[7]•1c ^{1}H-NMR

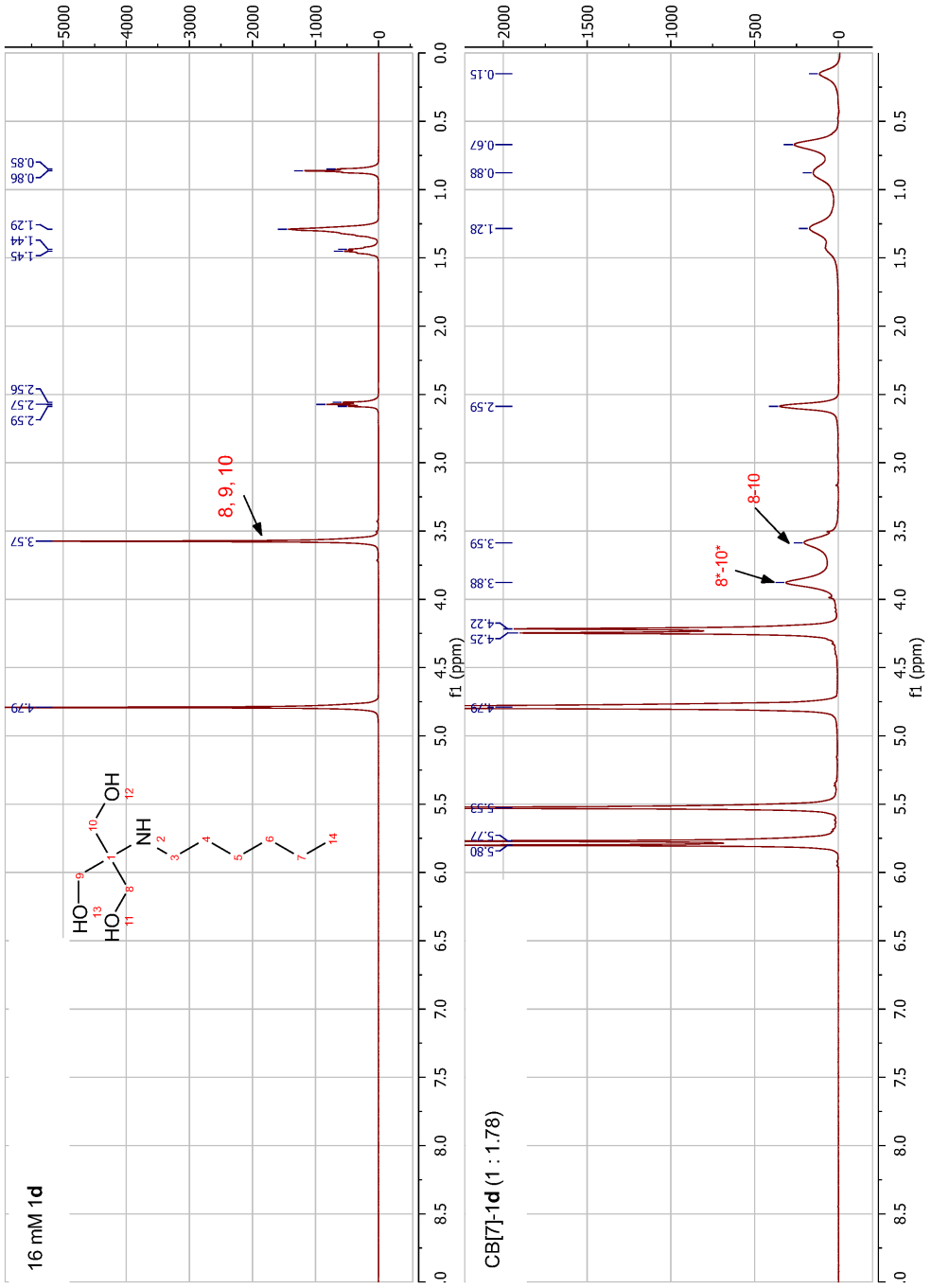


Figure 47: CB[7]•1d ^{1}H-NMR

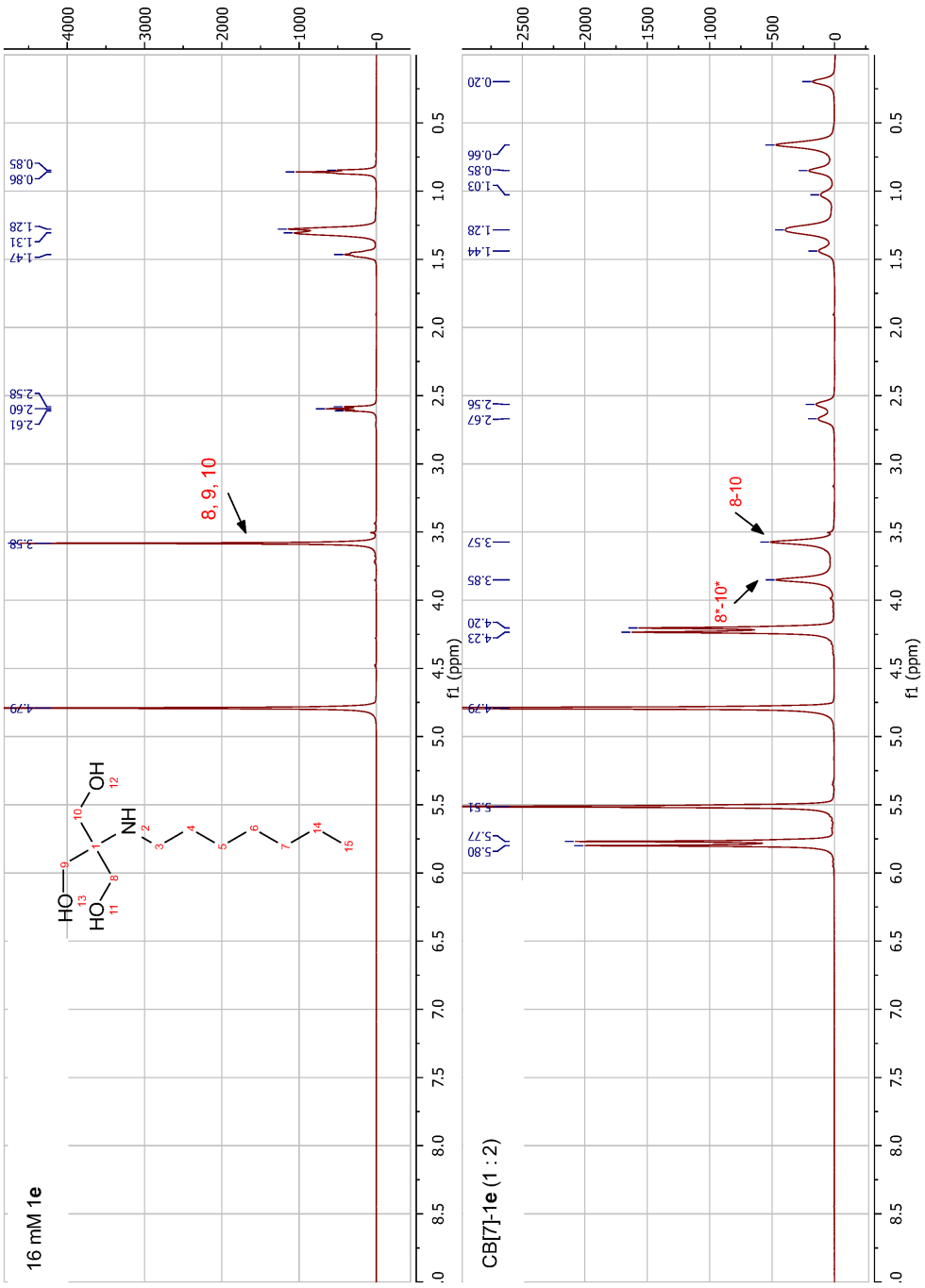


Figure 48: CB[7]•1e ^{1}H-NMR

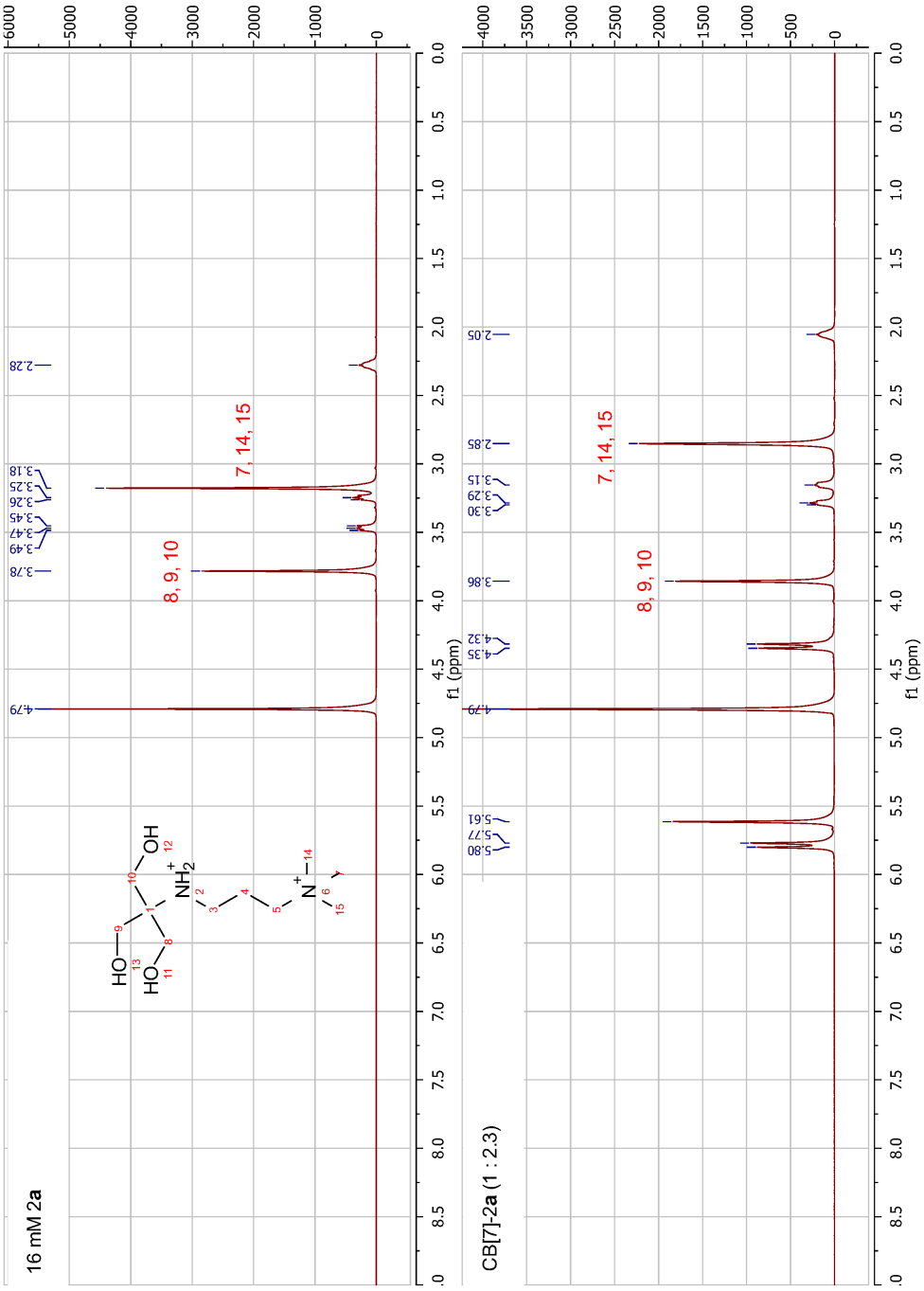


Figure 49: CB[7]•2a ^{1}H-NMR

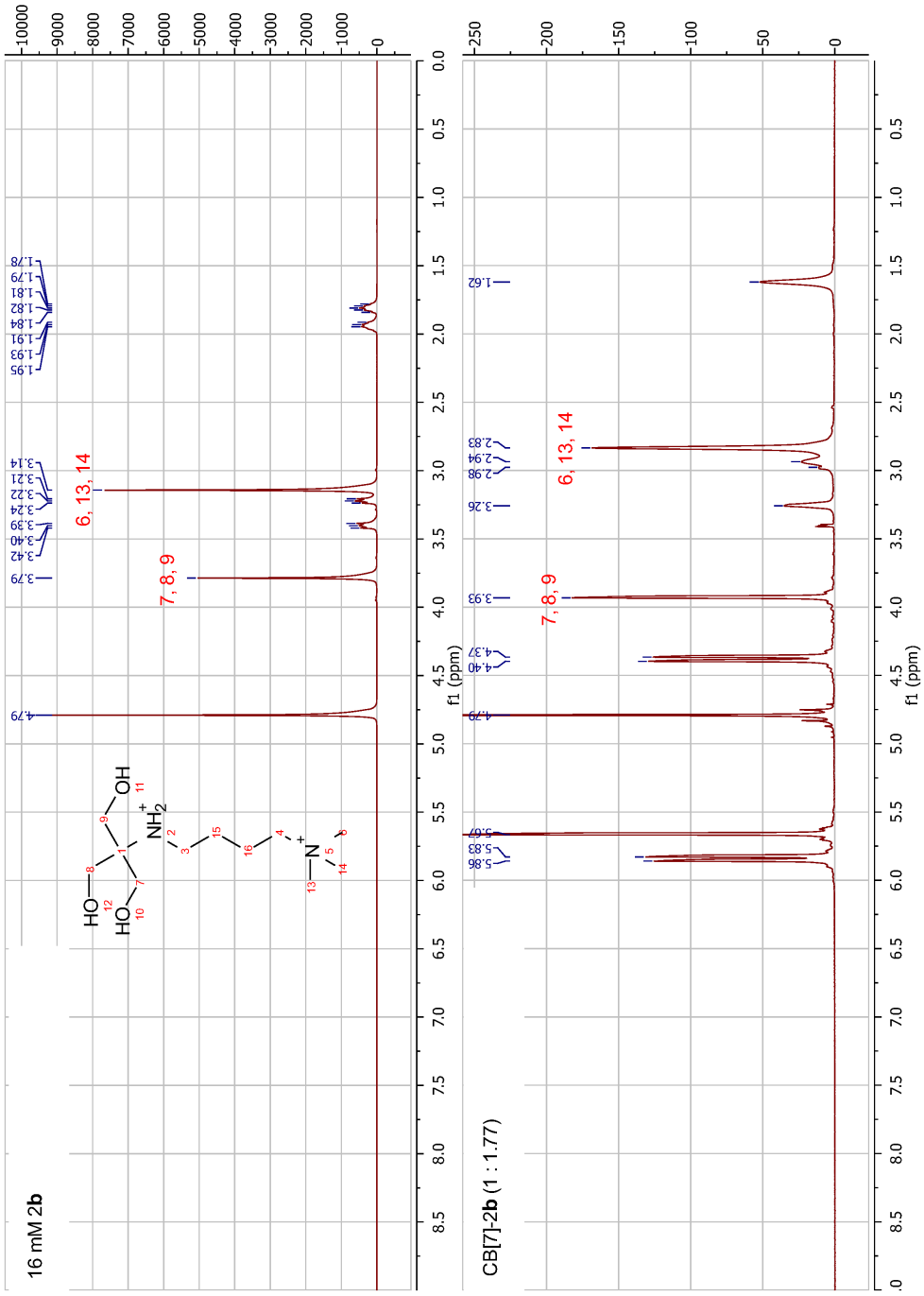


Figure 50: CB[7]•2b ^{1}H-NMR

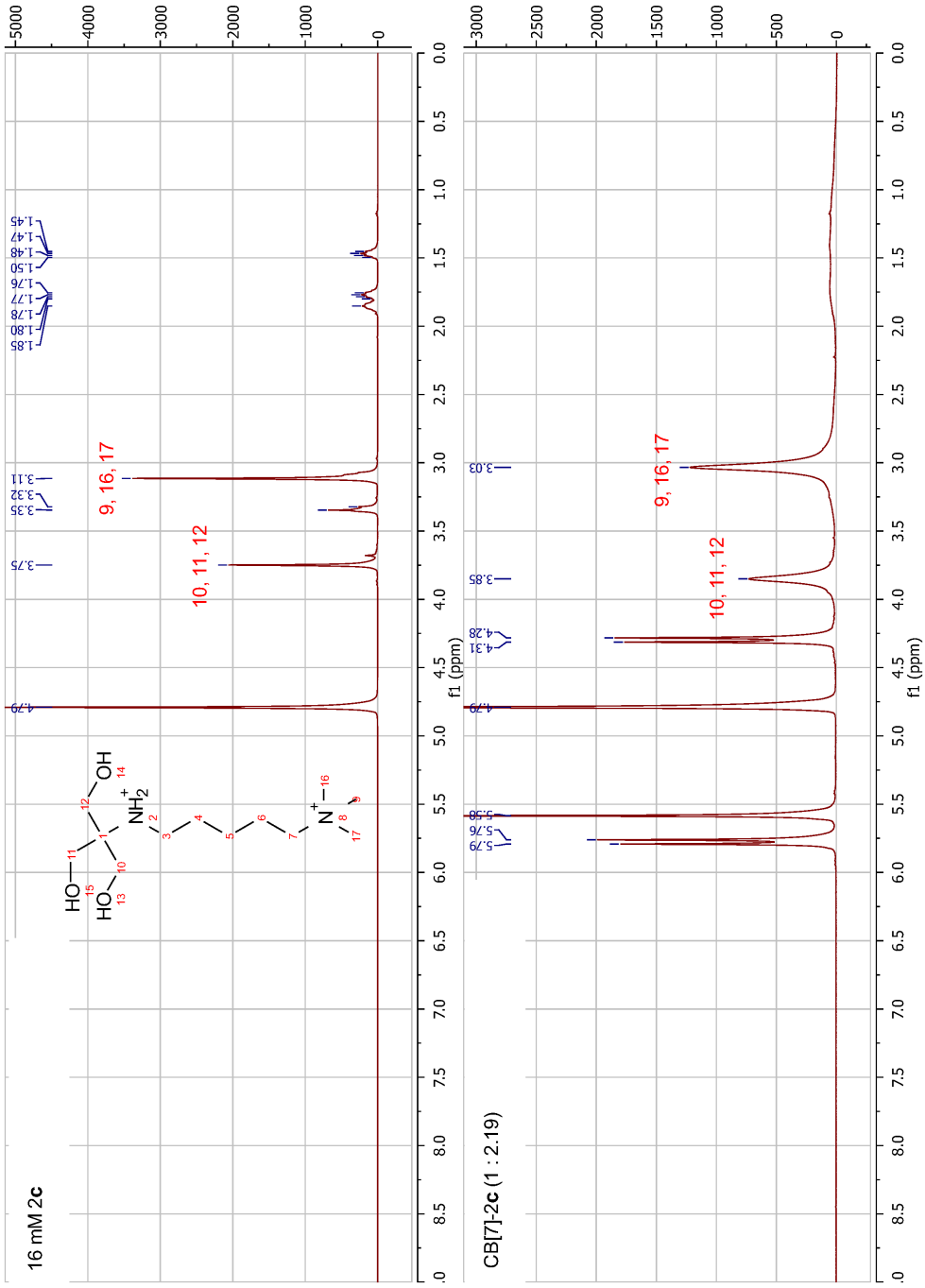


Figure 51: CB[7]•2c ^{1}H-NMR

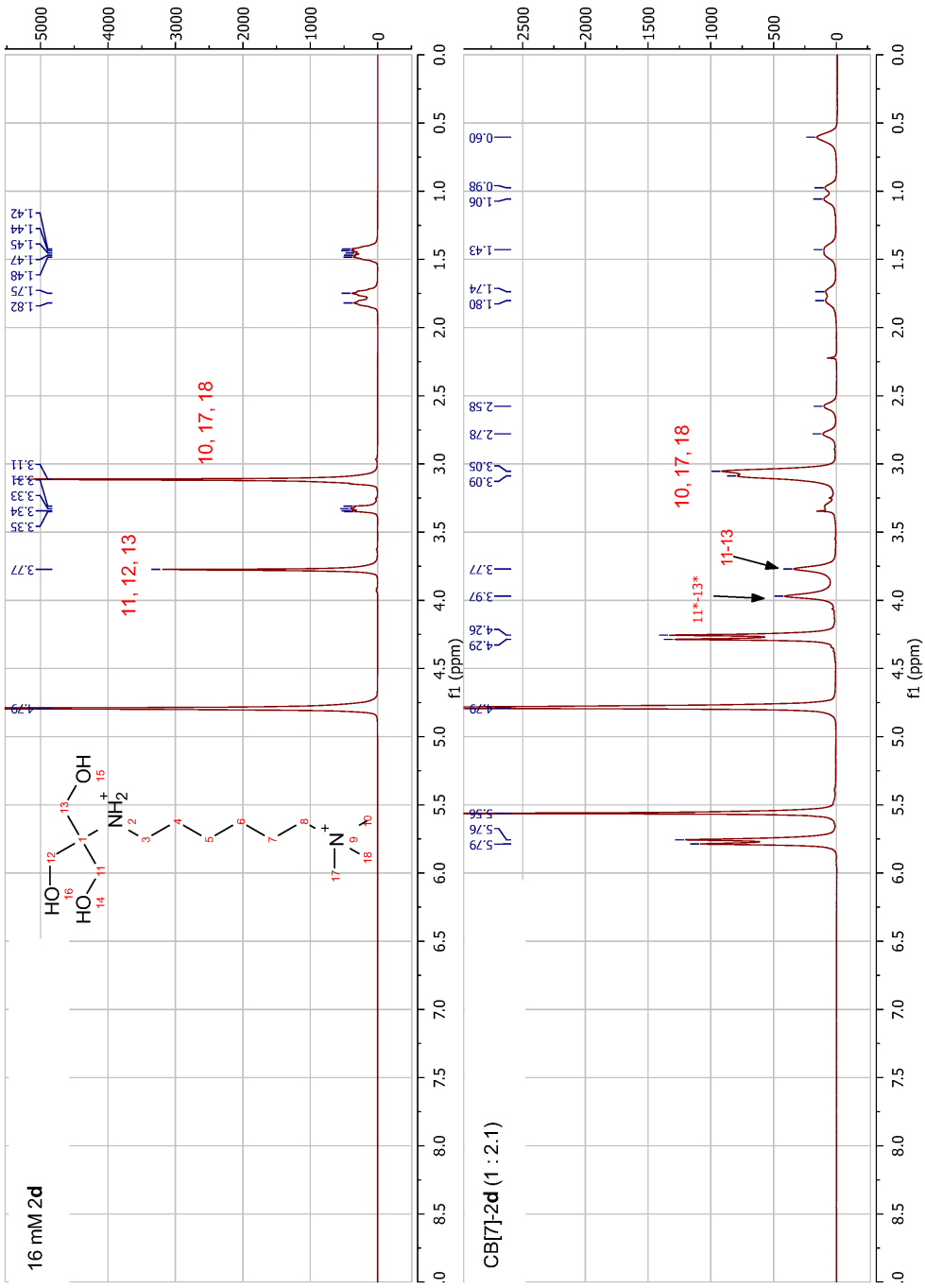


Figure 52: CB[7]•2d ^{1}H-NMR

of guests within the host. Ligand **1a** binds CB[7] in a structurally distinct fashion from the other members of the series with the Tris anchor encapsulated (Figure 44) and was removed from the study.

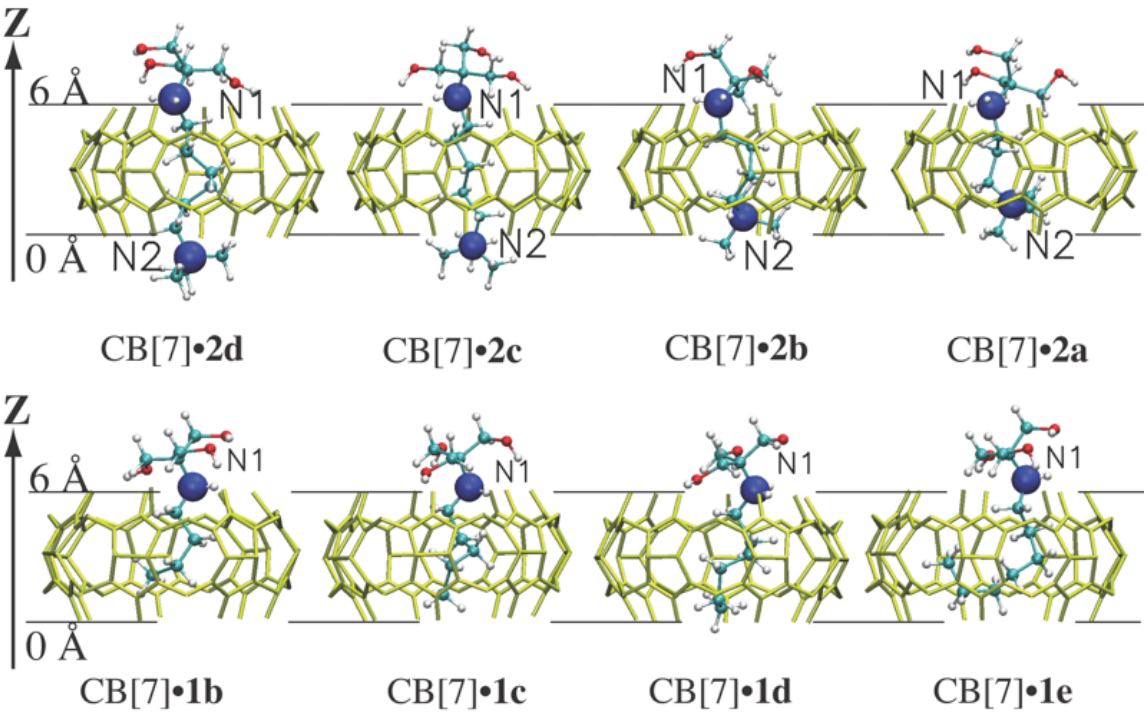


Figure 53: DFT structures of CB[7]•ligand complexes

To further verify the complex structures of our design, we computed energy minimized structures of the CB[7]•**2a-d** and CB[7]•**1b-e** complexes using QM/MM simulations with explicit solvent (Figure 53). The axis of rotational symmetry of CB[7] defines the Z-axis, and the seven lower portal oxygen atoms were placed at Z=0 Å. The computed offsets of N1 (Tris) and N2 (methonium, **2a-d**) along the Z-axis are shown in Table 7. N1 in all complexes is located above the upper portal of the CB[7] host with minimal Z-axis variation, in contrast to a significant change of N2 along the Z-axis

across the series **2d** to **2a**. The trends of the N1 and N2 Z-offsets agree well with chemical shift trends for Tris-methylene and methonium protons (Table 7).

Table 7: 1H-NMR chemical shift changes and DFT-calculated positions of Tris and methonium groups.

	Δδ (Tris) ppm	Z (N1) Å	Δδ (X*) ppm	Z (N2) Å
CB[7]•1a	0.06	N/D	-0.24	
CB[7]•1b	0.25	6.2	-0.61	
CB[7]•1c	0.29	6.3	N/A	
CB[7]•1d	0.29	6.3	-0.73	
CB[7]•1e	0.28	6.7	-0.65	
CB[7]•2a	0.19	6.1	-0.77	1.4
CB[7]•2b	0.25	6.2	-0.55	0.9
CB[7]•2c	0.22	6.6	-0.18	-0.6
CB[7]•2d	0.20	6.4	-0.04	-1.0

*X indicates the terminal methyl and methonium group protons. ND: not determined as ligand 1a binds in a unique fashion to the host. N/A: not applicable due to intermediate exchange kinetics on the NMR time scale.

We attribute the anchoring ability of Tris to the hydrophilicity of the hydroxymethyl groups, a notion supported by the study of the two similarly sized triethylamino and triethanolamino anchor groups.[213, 214] Hydrophilic anchor groups tend to remain in direct contact with the bulk water rather than to desolvate and enter the hydrophobic cavity of CB[7]. Experimental evidence from a series of negative controls

(i.e., other anchors that are internalized such as N-methylmorpholino and triethylamino anchors) in conjunction with excellent agreement between experimental and theoretical values of ligand positions within CB[7] provide strong support for the assertion that the position of the Tris anchor remains constant across the ligand series. This effect is visualized by plotting the calculated change in position (Z) vs. observed CB[7]-induced chemical shift perturbations for the Tris anchor and internalized methonium (Figure 54). Indeed with decreasing linker length, the methonium group (red squares) is gradually internalized—this is depicted in Figure 54 with Z = 0 Å at the lower portal of CB[7]—while the Tris group position (blue circles for **2a-d** and green triangles for **1b-e**) remains fixed in each ligand. The ability of the Tris anchor to lock guests in a fixed geometry serves as the foundation for the thermodynamic model.

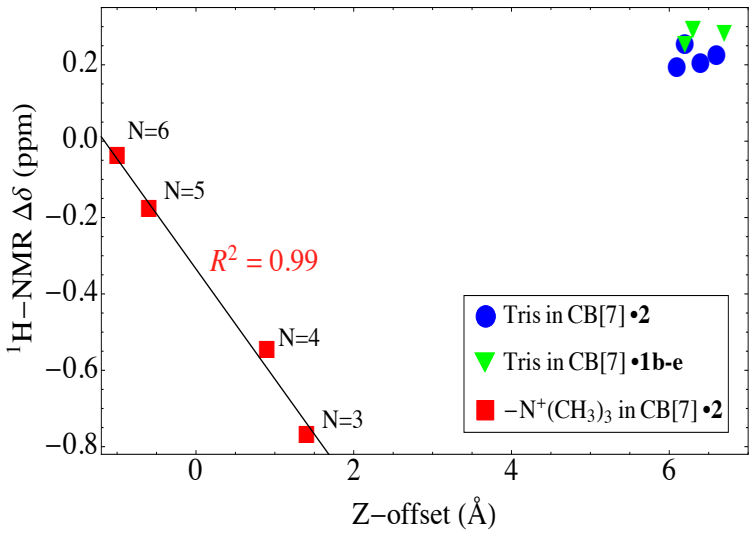


Figure 54: Correlation between experiment and theory depicts a gradual internalization of methonium with decreaseing linker length (N).

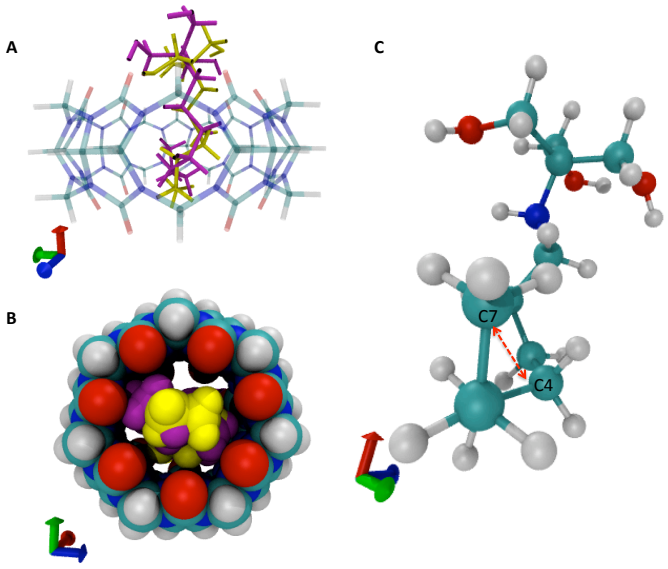


Figure 55: QM-optimized structures of 1d (pink) and 1e (yellow, C) bound to CB[7].

Results from ^{1}H-NMR studies suggest that the terminal methyl group of the CB[7]•**1** complexes remain inside the cavity regardless of linker length, an observation consistent with the energy-minimized structures. In the CB[7]•**1e** complex, the alkyl linker folds inside the cavity (Figure 48 and Figure 55), giving rise to a destabilizing C4-C7 *gauche* interaction (Figure 55c), an effect taken into account in the thermodynamic analyses below. In both the CB[7]•**2** and CB[7]•**1b-e** complexes, the alkyl linkers adopt helical conformations, similar to the geometries observed in cavitand-alkane complexes.[263] As previously described by Rebek and co-workers, the helical structure most likely results from the propensity of the alkyl chain to reduce solvent exposure and to fill the cavity volume as seen in the space filling model in Figure 55b.[264]

4.4 Deconvolution of methonium binding thermodynamics

4.4.1 Temperature variance in CB[7]•Ligand binding thermodynamics

We measured the CB[7]•**1b-e** and CB[7]•**2a-d** binding thermodynamics (ΔG, ΔH, and TΔS) over a range of temperatures from 208 – 313K (Table 8, Appendix D) with ITC. The binding is entropically driven with a favorable (negative) ΔH_b for **1c-e** and **2a-d** and an unfavorable $\Delta H_b^{CB[7]\bullet 1}$ for **1b**. ΔG_b does not appear to change with temperature due to compensating enthalpic and entropic contributions. Isobaric heat capacity changes accompanying CB[7] complexation were recorded for each ligand by plotting the binding enthalpy vs. temperature (Figure 56). Large, negative ΔC_p values were obtained for CB[7]•**1b-e** complexes, consistent with significant hydrophobic desovlation upon binding.

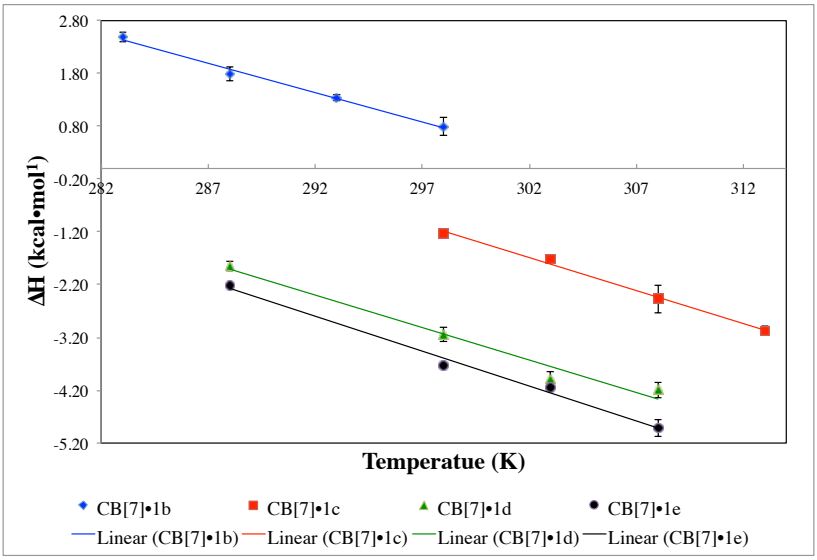


Figure 56: CB[7]•1b-e ΔC_p determined by ITC

Table 8: CB[7]•Ligand binding thermodynamics from ITC

Ligand	N	T (K)	ΔG (kcal•mol^{-1})	ΔH (kcal•mol^{-1})	TΔS (kcal•mol^{-1})	ΔCp (cal•mol^{-1}K^{-1})
CB[7]•**1b**	4	283	-5.28 ± 0.03	2.5 ± 0.10	7.78 ± 0.07	-109 ± 9
		288	-5.37 ± 0.08	1.79 ± 0.13	7.16 ± 0.09	
		293	-5.40 ± 0.06	1.34 ± 0.06	6.74 ± 0.02	
		298	-5.36 ± 0.11	0.79 ± 0.17	6.14 ± 0.17	
CB[7]•**1c**	5	298	-7.15 ± 0.05	-1.24 ± 0.03	5.92 ± 0.03	-116 ± 6
		303	-7.11 ± 0.03	-1.72 ± 0.05	5.39 ± 0.06	
		308	-7.17 ± 0.10	-2.47 ± 0.27	4.70 ± 0.19	
		313	-7.18 ± 0.05	-3.07 ± 0.09	4.10 ± 0.05	
CB[7]•**1d**	6	288	-8.03 ± 0.04	-1.85 ± 0.09	6.18 ± 0.06	-126 ± 7
		298	-8.04 ± 0.10	-3.14 ± 0.14	4.91 ± 0.06	
		303	-8.02 ± 0.04	-3.98 ± 0.13	4.04 ± 0.09	
		308	-7.91 ± 0.02	-4.19 ± 0.15	3.72 ± 0.14	
CB[7]•**1e**	7	288	-8.14 ± 0.04	-2.22 ± 0.02	5.91 ± 0.04	-140 ± 3
		298	-8.04 ± 0.04	-3.74 ± 0.04	4.30 ± 0.08	
		303	-8.07 ± 0.02	-4.14 ± 0.06	3.92 ± 0.06	
		308	-7.95 ± 0.05	-4.92 ± 0.16	3.03 ± 0.19	
CB[7]•**2a**	3	293	-6.83 ± 0.06	-0.37 ± 0.01	6.47 ± 0.08	-40 ± 3
		298	-6.80 ± 0.10	-0.61 ± 0.06	6.18 ± 0.05	
		303	-6.68 ± 0.04	-0.75 ± 0.02	5.93 ± 0.06	
		308	-6.73 ± 0.07	-0.99 ± 0.04	5.74 ± 0.11	
CB[7]•**2b**	4	288	-7.06 ± 0.15	0.37 ± 0.08	7.43 ± 0.11	-57 ± 3
		303	-7.40 ± 0.21	-0.44 ± 0.04	6.96 ± 0.24	
		308	-7.14 ± 0.47	-0.81 ± 0.04	6.33 ± 0.49	
		313	-7.33 ± 0.03	-1.04 ± 0.13	6.29 ± 0.13	
CB[7]•**2c**	5	288	-7.91 ± 0.02	-0.99 ± 0.03	6.93 ± 0.04	-57 ± 1
		298	-8.05 ± 0.03	-1.54 ± 0.01	6.52 ± 0.04	
		303	-8.00 ± 0.06	-1.83 ± 0.02	6.17 ± 0.07	
		308	-8.08 ± 0.11	-2.12 ± 0.11	5.95 ± 0.19	
CB[7]•**2d**	6	288	-9.02 ± 0.05	-1.69 ± 0.03	7.34 ± 0.03	-65 ± 6
		298	-9.08 ± 0.02	-2.42 ± 0.07	6.66 ± 0.08	
		303	-9.07 ± 0.02	-2.65 ± 0.09	6.43 ± 0.08	
		308	-9.08 ± 0.12	-2.97 ± 0.08	6.12 ± 0.16	

4.4.2 Determination of ΔJ_{CH_2} by ITC

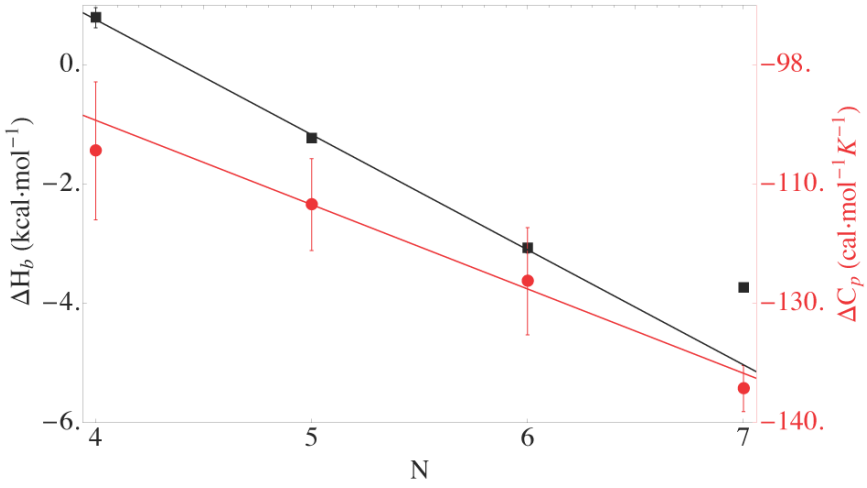


Figure 57: ΔH_b and ΔC_p vs. N for CB[7]•1b-e complexes

We treat each value of ΔJ_b as a linear function of N to estimate the per-methylene contribution (ΔJ_{CH_2}) of the linker to ΔJ_b. In accordance with the QM-optimized structure of the CB[7]•**1e** complex, a C4-C7 *gauche* interaction resulted in a 1.3 kcal•mol^{-1} deviation from the linear trend observed for ligands **1b-d** (Figure 57). Accordingly, we only use data for ligands **1b-d** to extract ΔH_{CH_2} (-1.9±0.1 kcal•mol^{-1}), ΔG_{CH_2} (-1.3±0.3 kcal•mol^{-1}) and ΔS_{CH_2} (-0.6±0.2 kcal•mol^{-1} at 298 K). As expected, the enthalpic cost of the solvent-independent *gauche* interaction does not translate into $\Delta C_p^{CB[7]\bullet 1}$, and a clear linear correlation with N is observed across the entire series **1b** to **1e** (Figure 57). The slope of this plot provides a per methylene contribution ($\Delta C_p^{CH_2}$) of -11±2 cal•mol^{-1}•K^{-1}, in good accord with previously reported values of -15 cal•mol^{-1}•K^{-1}.[265, 266] This correlation validates two critical assumptions: 1) heat capacity changes are

dominated by desolvation changes upon binding; and 2) the energetic contribution from desolvation of the alkyl linker varies linearly across the entire series **1b** to **1e**, regardless of the conformation adopted by the ligand in the bound form.

4.4.3 Enthalpy of desolvation of tethered methonium

Table 9 shows thermodynamic parameters for the binding of ligands **2a-d** ($\Delta J_b^{CB[7]\bullet 2a-d}$) at 298 K, which show that the binding of ligand series **2** to CB[7] is entropically driven. However, after correcting for the linker contribution (Equation 41), it is clear that methonium encapsulation is enthalpically driven (Table 9), in good accord with previous studies.[215, 217, 223, 235] The transfer of methonium from bulk water to the hydrophobic cavity of CB[7] is spontaneous ($\Delta G_{b,Am}^{6\to 3}$ = -1.7 kcal•mol^{-1}), driven by a negative (favorable) enthalpy change ($\Delta H_{b,Am}^{6\to 3}$ = -4.0 kcal•mol^{-1}), and accompanied by an unfavorable entropy change ($\Delta S_{b,Am}^{6\to 3}$ = -8 cal/mol•K). $\Delta\Delta H_{\mathrm{int},Am}^{6\to N}$ from MD simulations of

Table 9: CB[7]•2a-d and extracted CB[7]•methonium binding thermodynamics (kcal•mol^{-1}) at 298K

	N	ΔG_b	ΔH_b	$T\Delta S_b$	$\Delta\Delta G_{b,Am}^{6\to N}$	$\Delta\Delta H_{b,Am}^{6\to N}$	$T\Delta\Delta S_{b,Am}^{6\to N}$
CB[7]•**2a**	3	-6.8±0.1	-0.6±0.1	6.2±0.1	-1.7±1.0	-4.0±0.3	-2.3±0.9
CB[7]•**2b**	4	-7.2*	-0.1*	7.1*	-0.8±0.7	-1.5±0.2	-0.8±0.6
CB[7]•**2c**	5	-8.1±0.0	-1.5±0.0	6.5±0.0	-0.3±0.3	-1.0±0.1	-0.8±0.3
CB[7]•**2d**	6	-9.1±0.0	-2.4±0.1	6.7±0.1	0.0±0.2	0.0±0.1	0.0±0.1

*CB[7]•**2b** binding was enthalpically silent at 298K and the listed values were obtained by interpolation from other temperatures.

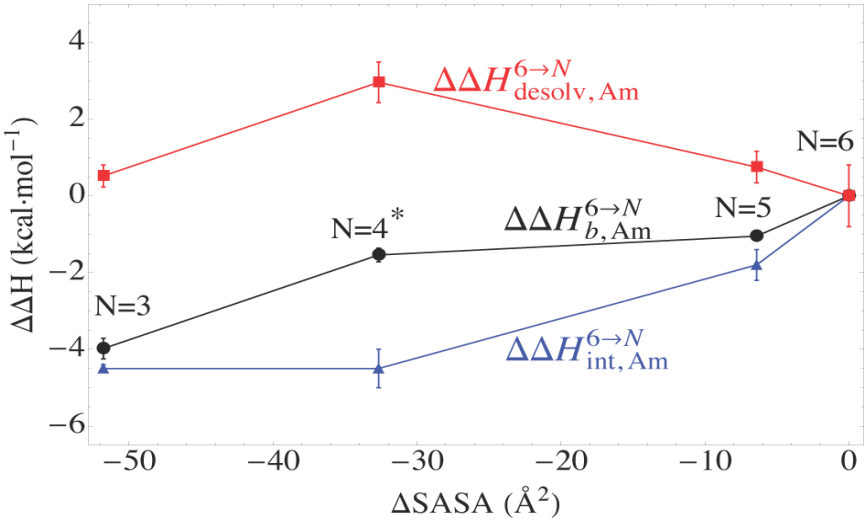


Figure 58: CB[7]•methonium binding enthalpy as function of QM calculated methonium solvent accessible surface area

the CB[7]•**2** complexes (Equation 63 of Appendix C) enable determination of the enthalpy of desolvation with respect to methonium encapsulation ($\Delta\Delta H^{6\to N}_{desolv,Am}$). Figure 58 shows the correlation between $\Delta\Delta H^{6\to N}_{b,Am}$ (black circles), $\Delta\Delta H^{6\to N}_{\text{int},Am}$ (blue triangles), and $\Delta\Delta H^{6\to N}_{desolv,Am}$ (red squares) as a function of the change in the ensemble averaged SASA relative to the N=6 state (ΔSASA). The binding of methonium to the CB[7] cavity produces a negative enthalpy change, presumably derived from favorable methonium-CB[7] interactions (N=6 to N=4). A large favorable change in binding enthalpy is observed as the linker is shortened from N=4 to N=3 ($\Delta\Delta H^{4\to 3}_{b,Am}$). Since the CB[7]-methonium interactions are similar in both ligands, this shift presumably arises from changes in desolvation enthalpy. The unfavorable net change in desolvation

($\Delta\Delta H^{6\rightarrow N}_{desolv,Am}$) correlates no more than weakly with the change in SASA, with a maximum observed at N=4. At least two explanations rationalize this observation.

First, energetic terms for which our thermodynamic model does not account, such as changes in the internal energy of the ligand or the receptor across the series, may manifest themselves in the derived values of $\Delta\Delta H^{6\rightarrow N}_{desolv,Am}$. Shortening of the linker from N=5 to N=4 begins to draw methonium into the cavity. The equilibrium diameters of the CB[7] portal and methonium are ~4-5 Å and ~5-6 Å, respectively.[267] Thus, at the equilibrium position of methonium in the CB[7]•**2b** (N=4) complex, the CB[7] portal may undergo distortion to accommodate the methonium group, or geometric constraints may force the methonium group to move further into or out of the cavity than would be predicted purely on the basis of linker length, incurring strain in the alkyl linker. Any of these effects would negate the simplifying assumptions of our thermodynamic model, producing additional destabilizing enthalpic contributions to the apparent $\Delta\Delta H^{6\rightarrow N}_{desolv,Am}$. At a minimum, structural distortions of the CB[7] cavity arising from methonium encapsultion were not apparent in a superposition of the QM calculated complex structures, depicting RMS differences no greater than 0.2 Å for CB[7] structures throughout the CB[7]•2 series.

A second more intriguing possibility is that $\Delta\Delta H^{6\rightarrow N}_{desolv,Am}$ does not vary monotonically with the solvent-exposed surface area, in which case the non-linearity in Figure 58 is representative of a more complex desolvation energy landscape. Similar

effects (*in silico*) have been discussed in the context of hydrophobic desolvation during helix formation[268] and the binding of a methane-sized particle in a hydrophobic cavity.[269, 270] Fortunately, these two possibilities are experimentally distinguishable: solvation-associated contributions to binding enthalpies are temperature dependent while the solute-solute interactions and configurational enthalpies are temperature independent.[271, 272] We can thus differentiate the possibilities by evaluating the changes in isobaric heat capacity that occur upon shortening of the linker and concomitant methonium encapsulation in CB[7] ($\Delta\Delta C_{p,Am}^{6\rightarrow N}$).

4.5 The role of water in CB[7]•methonium binding

4.5.1 Heat capacity as a measure of solvent reorganization

Figure 59 shows the correlation of $\Delta\Delta C_{p,Am}^{6\rightarrow N}$ and $\Delta\Delta H_{desolv,Am}^{6\rightarrow N}$ with ΔSASA. Similar to $\Delta\Delta H_{desolv,Am}^{6\rightarrow N}$, $\Delta\Delta C_{p,Am}^{6\rightarrow N}$ does not follow a simple monotonic trend but rather shows an apparent minimum at N=4. Propagation and accumulation of errors in $\Delta\Delta C_{p,Am}^{6\rightarrow N}$ preclude precise assessment of the heat capacity changes associated with methonium transfer into the cavity of CB[7]. Nonetheless, a discontinuity in $\Delta\Delta C_{p,Am}^{6\rightarrow N}$ at N=4, qualitatively coincides with the maximum observed in $\Delta\Delta H_{desolv,Am}^{6\rightarrow N}$ prior to complete methonium encapsulation. In light of the absence of significant distortion of CB[7] for N=4 (*vide supra*), the observed $\Delta\Delta C_{p,Am}^{6\rightarrow N}$ supports the notion that an unfavorable enthalpic component for N=4 arises from solvent reorganization, rather

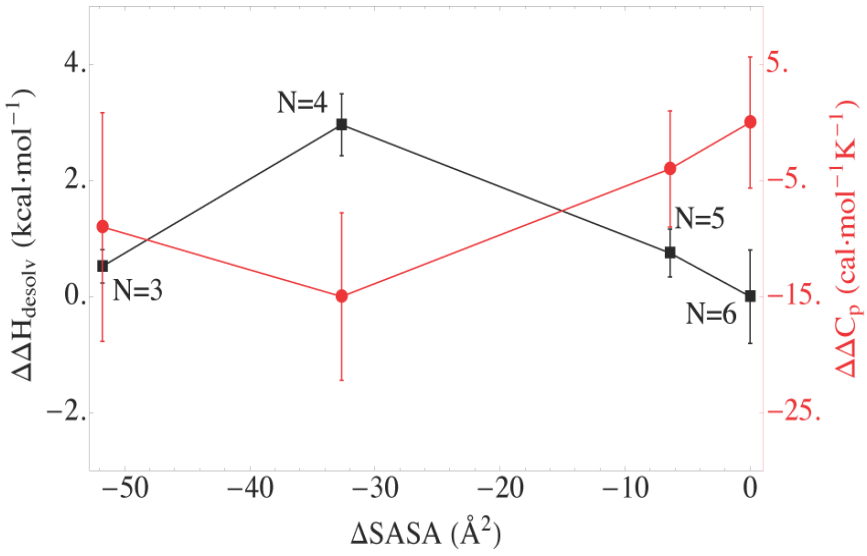


Figure 59: Comparison of methonium desolvation enthalpy and CB[7]•methonium binding heat capacity changes as a function of ΔSASA

than from increased torsional strain in the CB[7]•**2** complexes. Also consistent with this hypothesis are earlier computational studies that predicted a non-monotonic trend of desolvation enthalpy with decreasing intermolecular distance upon the association of non-polar surfaces in water.[273]

The non-monotonic nature of the $\Delta\Delta H^{6\rightarrow N}_{desolv,Am}$ and $\Delta\Delta C^{6\rightarrow N}_{p,Am}$ plots prevents unambiguous determination of the desolvation enthalpy of a fully solvated methonium by linear extrapolation (SASA = 167 Å²). However, given an exceptionally small enthalpy change in the partial desolvation of methonium ($\Delta\Delta H^{6\rightarrow 3}_{desolv,Am}$ = 0.5±0.3 kcal•mol^{-1}), it appears highly unlikely that the enthalpic penalty for transfer of the full methonium surface from water to a hydrophobic cavity will be close to the gas-phase value for TMA$^+$ desolvation enthalpy (+49 kcal•mol^{-1}). On the other hand, our results are

not necessarily inconsistent with TMA^+ gas-phase data. An intriguing explanation for this discrepancy is that methonium retains significant stabilizing interactions with bulk water even when encapsulated in the CB[7] pore. Even upon full encapsulation (N=3), methonium is only separated from the nearest water molecules by the 6-8 Å wall of CB[7] cavity. The low dielectric constant of the CB[7] structure may enable strong methonium-water electrostatic interactions. In contrast to gas-phase studies, where the water-TMA^+ electrostatic interactions disappear upon removal of solute, these interactions may be substantially retained even after the transfer of methonium from water to the CB[7] cavity, giving rise to the diminished enthalpic penalty of desolvation.

A review of methonium binding proteins shows that methonium is typically removed from bulk water by less than 20 Å, albeit in a pocket surrounded by aromatic moieties.[215-219, 274] The enthalpic behavior of methonium desolvation observed here may reflect Nature's strategy for methonium binding in proteins – although methonium is *desolvated* in the sense that it lacks direct *physical* contact with water, the charge it carries remains in *electrostatic* contact with solvent, a situation starkly different from that in gas phase studies. As a result, the large enthalpic penalty found in gas phase studies may be mitigated in the binding of methonium to CB[7] and other receptors, because of indirect interactions with solvent. This postulate warrants further testing.

4.5.2 Occupancy maps of solvation shell water of methonium

To further understand the non-linearity in our energy-surface plots, we calculated local occupancy maps for water surrounding methonium in the CB[7]•**2** complexes (Figure 60) based on the MD simulation trajectories. The first solvation shell of the solvated states has a water occupancy higher than that of the bulk. Collapse of this first solvation shell takes place well before methonium enters the cavity (from N=5 to N=4), an observation consistent with the results of McCammon and others.[268-270] At N=4, the first solvation shell of methonium nearly vanishes, *prior* to full encapsulation at N=3. Disruption of the solvation shell upon binding was shown to be unfavorable enthalpically[269, 270, 275] and may explain the observed trend in desolvation thermodynamic parameters: the low-occupancy region near methonium in the CB[7]•**2b** complex contains energetically perturbed waters (relative to the bulk) that produce both the enthalpic discontinuity at N=4 and the minimum in $\Delta\Delta C_{p,Am}^{6\rightarrow N}$. The release of this so-called activated water[276] to the bulk liberates $-\Delta\Delta H_{desolv,Am}^{4\rightarrow 3} \approx 2.5$ kcal•mol^{-1} at 298 K. Although the enthalpic maximum at N=4 may be due to the unique curvature of the surface constituted by the CB[7] portal and methonium,[277, 278] the observation of perturbed water is common in both biological and abiological systems.[275, 279, 280] At the very least, the presence of a complex desolvation energy landscape for the CB[7]•**2b** complex calls into question the validity of using simple water to either gas phase or

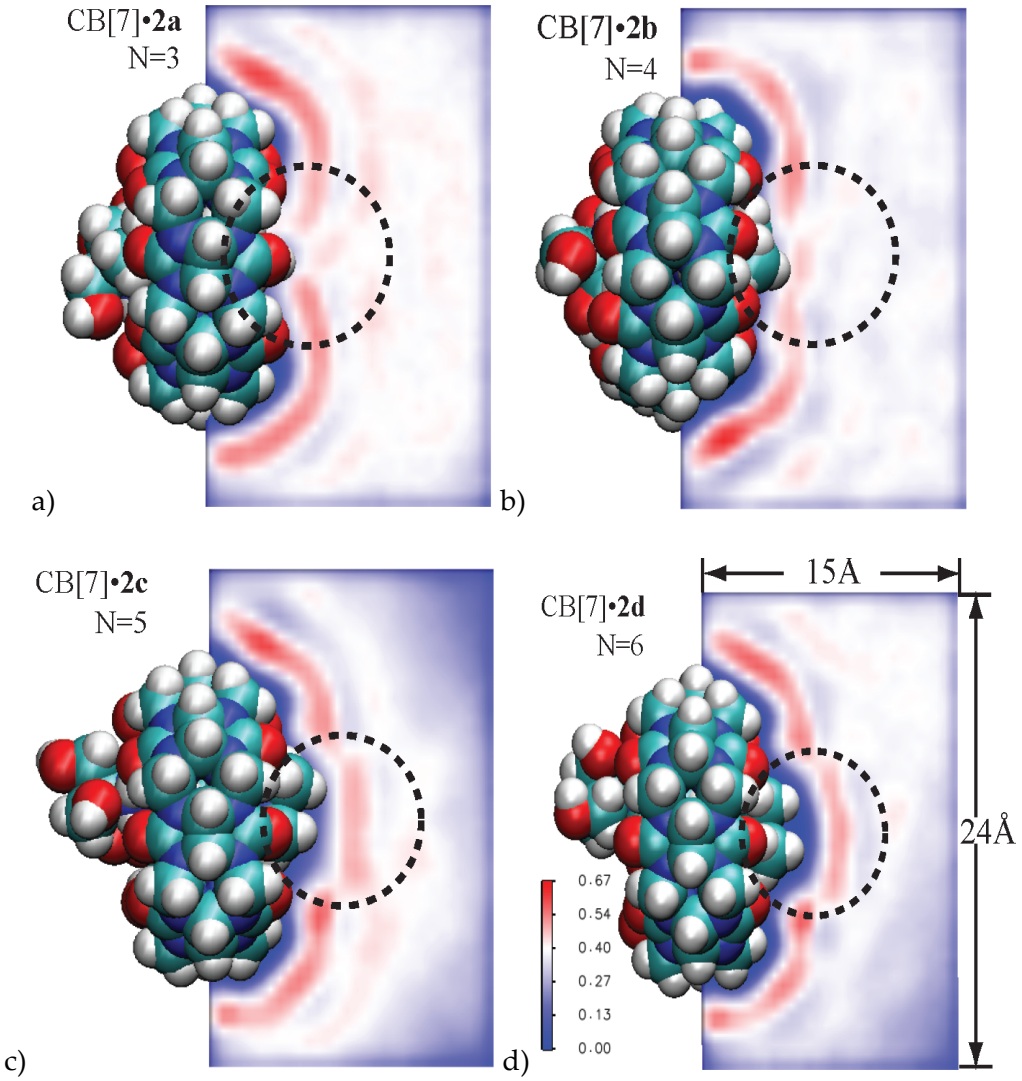


Figure 60: Water occupancy maps from MD simulation depicting solvation shell around solvent-exposed methonium in CB[7]•2a-d complexes

octanol partition coefficients to assign the energetic contributions of water for ligand binding in aqueous solution.

4.6 Conclusions

We used a unique host-guest system comprising CB[7] and a series of *de novo* designed ligands to determine the thermodynamics of methonium desolvation and

binding to a hydrophobic cavity. In this system, a Tris anchor fixes a ligand in place with respect to the host, and methonium is incrementally repositioned from bulk water into the cavity by shortening an alkyl linker. The transfer of methonium from the partially solvated state of the N=6 ligand to the fully encapsulated state at N=3 is driven by exothermic methonium-CB[7] interactions. The data presented here furnish evidence that the desolvation of methonium occurs with a significantly diminished enthalpic penalty relative to the gas-phase desolvation of TMA^+. The small values of heat capacity changes that accompany binding suggest minimal solvent reorganization upon encapsulation of methonium, consistent with neutron scattering studies that show a remarkably unperturbed solvation shell around the methonium ion.[233]

More importantly, our approach of incremental internalization revealed non-monotonic trends in both $\Delta\Delta C_{p,Am}^{6\rightarrow N}$ and $\Delta\Delta H_{desolv,Am}^{6\rightarrow N}$ as a function of methonium SASA, which echoes the arguments of Lemieux, Dill and others that not all surface is created equal and approaches to parameterize desolvation thermodynamics based solely on SASA have serious limitations.[277-279] For a surface as simple as that of methonium, the enthalpic signature of binding desolvation depends strongly on the molecular topology of the surface formed by both the ligand and the receptor. The thermodynamic consequences of methonium binding to different receptor surfaces may be different despite the fact that equal SASA is desolvated. Our result reflects the common limitation of SASA-based thermodynamic analysis – namely that they operate well across a homologous series but

fail on transfer to settings other than that from which the parameters are generated. As such, new methods of capturing relevant desolvation thermodynamic parameters for association in aqueous solution are needed for more accurate modeling of aqueous desolvation, a notion recently articulated by Baldwin.[281]

In any event, spontaneity of desolvating methonium from water is determined by the free energy of transfer into the cavity of CB[7]. The thermodynamic model developed here for enthalpy may not apply to free energy analysis, as the additive partitioning of $\Delta S_b^{CB[7]\bullet 2a-d}$ may not be valid. Gas phase measurements of binding free energies and enthalpies in CB[n] systems offer an attractive means for establishing a free energy description of methonium desolvation, which was recently reviewed by Yang and Dearden.[282] We are currently exploring this approach.

Finally, we note that the study described here accesses only half of the total SASA of methonium. Indeed the SASA of a fully solvated methonium group is 167 Å^2, while the most solvent-exposed state of methonium in the CB[7]•**2** complexes (N=6) presents roughly half of this area to solvent. The correlation plots of $\Delta\Delta C_{p,Am}^{6\rightarrow N}$ and $\Delta\Delta H_{desolv,Am}^{6\rightarrow N}$ may represent asymptotic regions of two much steeper curves, and the desolvation of the entire 167 Å^2 SASA of methonium may entail a much greater enthalpic penalty than suggested by our data. Due to the flexible nature of the alkyl linker, however, the current design is unsuitable for probing more solvent exposed states of methonium. Future studies using alkynyl or alkenyl linkers to extend the

methonium further into the solvent may access a wider range of solvent-accessible states of methonium.

5. Experimental section.

5.1 General

5.1.1 Materials

House deionized water was further purified prior to use with a Milli-Q Advantage A10 system fitted with an in-line 0.25 μm filter (EMD-Millipore, Billerica, MA, USA). *E. coli* strains XL10(Gold) or DH5α (Stratagene, La Jolla, CA, USA) were used for general cloning. Top10 One Shot® Cells (Invitrogen, Carlsbad, CA, USA) were used for subcloning of PCR products for sequencing purposes, and BL21(DE3) (Novagen, Madison, WI, USA) cells were used for protein expression under control of the lac repressor/ T7 RNA polymerase expression system. All DNA-altering enzymes were purchased from New England Biolabs (NEB, Ipswich, MA, USA). Nano-Strip® was purchased from Cyantek (Fremont, CA, USA). TCEP and IPTG were purchased from GoldBio Technology (St. Louis, MO, USA). P2-Biogel and Kaleidoscope® protein ladder were purchased from Bio-Rad (Hercules, CA, USA) and used as per manufacturer's instructions. HR-S100 gel filtration column was purchased from GE Healthcare (Pittsburgh, PA, USA). Pivalic acid was purchased from TCI-America (Portland, OR, USA). 4Å molecular sieves (4ÅMS; Sigma-Aldrich, St. Louis, MO, USA) were flame dried for 10 min. at 150 mTorr and stored in an oven at 200 °C. Di-tert-butyl-4-methylpyridine (DTBMP) was ordered from Matrix Scientific (Columbia, SC, USA) and purified by elution over 1 volume SiO_2 and 4 volumes basic alumina in hexanes, yielding

a white solid after crystallization from the pure concentrate. Guanidine was generated as a free base by titration of the hydrochloride solution (Sigma-Aldrich) with NaOH and desalting over mixed DOWEX-H/OH resin then freeze-dried to a white powder. Silver (I) trifluoromethanesulfonate (AgOTf) was purchased from Sigma-Aldrich and azeotropically dried in triplicate with toluene prior to use. Benzaldehyde (PhCHO), 4-(dimethylamino)pyridine (DMAP), N-bromosuccinimide (NBS), copper(II) trifluoromethanesulfonate ($Cu(OTf)_2$), and trimethylacetyl chloride (PivCl) were purchased from Sigma-Aldrich and re-purified according to literature procedures.[283] All other chemicals were purchased and used without modification from Sigma-Aldrich.

5.1.2 Instrumentation

XPS spectra were collected on a Kratos Analytical Axis Ultra spectrometer (Spring Valley, NY, USA) and analyzed using the CasaXPS platform (Kratos Analytical, Teignmouth, Devon, ENG). Cellulose film thickness was measured using the Nanometrics 210 reflectometer (Kanata, ON, CAN). Contact angles were collected on an NRA C. A. model 100 contact angle goniometer (Rame-Hart Inc., Succassuna, NJ, USA). AFM images were collected on a Digital Instruments Dimension 3100 scanning probe microscope (Veeco Metrology, Santa Barbara, CA, USA) and rendered using Nanoscope SPM software from Veeco. UV/Vis spectroscopy was performed using the Hewlett-Packard model 8453 spectrometer (Agilent Technologies, Cary, NC, USA). Micro-plate UV/Vis assays were performed on a Spectramax Plus 384 microplate reader and the data

analyzed by SoftMax Pro version 5.0.1 software (Molecular Devices, Sunnyvale, CA, USA). NMR spectra were collected on an Agilent DD2 800 (Agilent Technologies), Varian Inova 500, Varian Inova 400, or Varian Inova 300 (Varian Inc., Palo Alto, CA, USA). ^{1}H-NMR chemical shifts (δ) are reported in reference to TMS (0.00 ppm). ^{1}H-NMR peak multiplicities are reported as singlet (s), doublet (d), triplet (t), doublet of doublet (dd), doublet of triplet (dt), AB quartet (ABq), and multiplet (m) with coupling constants (*J*) reported in Hz. ^{13}C-NMR chemical shifts are reported in reference to internal solvent. Liquid-chromotography/ mass spectrometry (LCMS) spectra were collected using an Agilent Ion Trap with electrospray ionization and ion collection in positive mode (ESI+). Matrix-assisted laser desorption ionization- time of flight (MALDI-TOF+) spectra were collected on an ABI Voyager DE Pro (Life Technologies, Carlsbad, CA, USA) in positive/ linear mode. Hydronium concentrations were measured with a SympHony model SB70P pH meter (VWR, Radnor, PA, USA) fitted with a Calomel pHree glass electrode (Beckman Coulter, Brea, CA, USA) and calibrated to two or three pH standards (BDH brand, pH = 4.01, 7.00, and 10.01; VWR). Masses were measured to ± 1% using a Mettler-Toledo model AB104-S or AB204-S analytical balance (Mettler-Toledo, Columbus, OH, USA).

5.1.3 Synthesis and chromatography

All reactions were run under inert gas atmosphere using N_2 or Ar, unless specified otherwise. Hydrogenation reactions were performed with a Parr model 3911

hydrogenation apparatus (Parr Instrument Company, Moline, IL, USA). Solvent removal was facilitated with a Büchi model R110 rotary evaporator (Buchi Corp., New Castle, DE, USA). Lyophilization was performed with a Labconco Freezone 6 freeze-dry system (Labconco, Kansas City, MO, USA). Routine sample sonication was performed using an Aquasonic model 75T sonication apparatus from VWR. Normal phase thin-layer chromatography (TLC) was performed over glass plates coated with Silica-60F (254 nm) in a specified solvent then developed with phosphomolybdic acid solution or 1% (w/v) $KMnO_4$ in aqueous hydroxide, or visualized by UV (254 nm). Sugars were visualized using a 0.2% ethanolic naphthorescorcine (50% in 2N H_2SO_4) stain or 0.3% (v/v) *p*-anisaldehyde stain in acidic (9.5 mL conc. H_2SO_4 + 2.7 mL acetic acid) ethanol (250 mL) and charring technique. Reverse-phase TLC was performed with glass C18-modified silica gel plates (Silica $R18_F$, EMD-Millipore). Flash chromatography was performed over SiO_2 RediSep columns using a Teledyne-Isco Combi-Flash instrument (Teledyne-Isco, Lincoln, NE, USA). Normal phase analytical high pressure liquid chromatography (NP-HPLC) was performed using an Agilent Carbohydrate Column (150 mm x 2.5 mm ID; 25% - 100% H_2O: ACN; 60 °C) with elution monitored by MS. Reversed phase HPLC (RP-HPLC) was performed using an analytical and preparatory scale Agilent Prep-C18 column (250 mm x 2.5 mm ID; 30 cm x 4.5 cm ID) with analytical scale elution monitored by MS. Semi-preparative HPLC was performed on an Agilent series 1200 purification system using the Agilent Chemstation software and monitored via an in-line UV/Vis

detector. Stringent FPLC was performed using an ÄKTAFPLC system from GE Healthcare with an in-line UV/Vis detector. Routine FPLC was also performed using a Masterflex® pump (Cole-Parmer, Court Vernon Hills, IL, USA) to control solvent flow and a Pharmacia model LKB RediFrac fraction collector of GE Healthcare for sample collection.

5.1.4 Biochemistry and molecular biology

All containers and media were either acquired in sterile-packaging or autoclaved using the AMSCO Scientific Eagle/ Century model SG-120 autoclave (Steris, Mentor, OH, USA). DNA reactions were performed with an Applied Biosystems GeneAmp® PCR system 2700 from Life Technologies. DNA sequencing was performed by the Duke University core facilities (Durham, NC, USA). Cell transformations were performed using a Bio-Rad MicroPulser™ on setting E2 for electrocompetent cells[284] or by the heat-shock method for chemically competent cells.[285] Prior to plating, transformants were generally incubated in SOC media for 45 min. while stirring in a New Brunswick Scienitic model TC-7 rotor housed within a VWR model 1545 oven thermostated at 37 °C. Centrifugation of small volumes (< 2 mL) was performed with a Beckman Coulter model Microfuge® 18 fixed-angle centrifuge. Routine centrifugation and Centricon (EMD-Millipore) protein concentration were performed on an Eppendorf model 5810-R bench-top swing-rotor centrifuge (Eppendorf, Hamburg, DEU). Large volume centrifugations were performed using a Sorvall RC5B Plus centrifuge with a GS-3 or SS-

34 fixed-angle rotor (DuPont, Research Triangle Park, NC, USA). Ultracentrifugation was performed using a Beckman Coulter Optima LE-80K Ultracentrifuge with a type 45-Ti fixed-angle rotor. Cell cultures were grown in beveled shaker flasks within thermostated shaker-incubtors: Infors-HT model Multitron I shakers (Infors USA Inc., Laurel, MD, USA) at T ≥ 25 °C; a Thermo Scientific model MaxQ 5000 shaker (Thermo Fisher Scientific, Waltham, MA, USA) for T = 20 - 25 °C; and a New Brunswick Scientific series 25 incubator shaker (New Brunswick Scientific Company Inc., Edison, NJ, USA) housed within a cold environment for T = 4 - 20 °C. Cell disruption was performed using either an EmulsiFlex-C5 homogenizer (Avestin Inc., Ottawa, Ontario, Canada) or an SLM Aminco® French press (SLM Instruments, Ivyland, PA, USA). Gel electrophoresis was performed using the Bio-Rad Power Pac 3000 set to 200V for SDS-PAGE and 60 – 80 V for agarose gel electrophoresis. PAGE gels were stained with 0.25 wt% Brilliant Blue Coomassie stain and subsequently de-stained with 10% acetic acid. Gel documentation was performed with a Kodak Gel Logic Imaging system using the Molecular Imaging Software platform (Eastman Kodak, Rochester, NY, USA).

5.1.5 Isothermal Titration Calorimetry

ITC molecular association data was collected on a VP-ITC model calorimeter using Origin 5 data analysis software (GE Healthcare Life Sciences). A typical ITC titration was carried out by titrating 30 aliquots of ligand solution (15 μL) into a sample solution (typically the target receptor protein or organic host), with 350 s intervals

between injections. Ligand concentrations were ~12-15-fold greater than that of the sample solutions, which in turn were set to ensure *c*-values remained between 1 and 1000 (see Equation 31). Chapter 4 ligand and CB[7] concentrations were determined gravimetrically assuming molecular formulas obtained by combustion analysis. Cellodextrin ligand concentrations were determined from using the BCA reducing cellodextrin assay provided below. Protein concentrations were determined using the Edelhoch method.[286] Reaction heats were typically fit to a one-sites binding polynomial (Equation 30) to determine the association constants and binding thermondynamics. Representative CBM•cellodextrin titrations are shown in chapter 2, and CB[7]•ligand titrations are provided in Appendix D. CB[7]•ligand titrations were performed in 100 mM NaAcetate buffer, pH = 4.6. CBM3a•cellodextrin titrations were performed in 50 mM NaHEPES, 1 mM $CaCl_2$, 1 mM 2-mercaptoethanol, pH = 7.5.

5.1.6 General protocols for protein expression and quantitation

IPTG-induced gene expression. BL21(DE3) *E. coli* were transformed with pET-based expression plasmids using electroporation[284] or the heat shock method.[285] Single colonies were used to inoculate a 5 mL starter culture in LB broth supplemented with 50 mg•L^{-1} of appropriate antibiotic. The starter culture was rotated overnight in a 37 °C incubator. The starter culture was used to inoculate a 0.5 L culture of LB broth containing appropriate antibiotic. The culture was grown at 37 °C in beveled flasks with shaking at 200 rpm until the optical density measured at 600 nm reached an absorbance of 0.4 – 0.6.

The cells were cooled to the appropriate temperature (typically 15 – 37 °C) and a concentrated solution of IPTG in water (500X) was added to the culture. Small aliquots (6 mL) of the induced culture were periodically removed and the cells collected by centrifugation at 4,000 rpm and 4 °C. The cell pellets were subjected to alkaline lysis, and the total protein content measured by standard protocols in reference to bovine serum albumin as described below. At the end of the expression time-course, the remaining cell slurry (~450 mL) was centrifuged at 5,000 rpm for 10 min and 4 °C. The cell pellet was subsequently stored at -20 °C prior to small scale lysis under denaturing or native conditions as described below.

Auto-induction of gene expression with lactose. An alternative method for protein expression was explored based on the method reported by Studier.[206] The following solutions were prepared:

1000X Trace Minerals solution: 50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM ZnSO4, 2 mM $CoCl_2$, 2 mM $CuCl_2$, 2 mM $NiCl_2$, 2 mM Na_2MoO_4, 2 mM H_3BO_3. Components usually combined to a final volume of 50 mL with Milli-Q water then sterile filtered and stored at 4 °C under foil.

25XN solution: 1.25 M Na_2HPO_4, 1.25 M KH_2PO_4, 1.25 M NH_4Cl, 125 mM Na_2SO_4. The components were combined in Milli-Q water up to 200 mL and autoclaved at 121 °C for 15 min. The resulting pH was 7.4.

50X5052 solution: 25% (v/v) glycerol, 2.5% (wt/v) D-glucose (anhydrous), 10% (wt/v) D-lactose monohydrate. The components were combined and diluted to 100 mL with Milli-Q water then sterile filtered and stored at 4 °C.

The auto-induction media was prepared as follows. For 0.5 L media, tryptone (5 g) and yeast extract (2.5 g) were combined with Milli-Q water (468.5 mL) and autoclaved at 121 °C for 15 min. 25XN solution (20 mL), 50X5052 solution (10 mL), 1 M $MgSO_4$ (1 mL; sterile filtered), and 1000X trace minerals solution (0.5 mL) was added when cool, and the media was supplemented with appropriate antibiotic. A starter culture of pET-transformed BL21(DE3) *E. coli* in LB/antiobiotic broth (*vide supra*) was used to inoculate the auto-induction media. The culture was grown at various temperatures (typically 15 – 30 °C) for 24 – 48 hours while shaking at 200 rpm in beveled flasks. Care was taken not to exceed 25% of the flask volume to ensure sufficient aeration. Small intermediate aliquots and the end-point cell pellet were isolated and treated for protein expression testing as described for the IPTG-induced cells.

Analytical-scale cell lysis with alkaline buffers and protein quantitation. Buffers P1, P2, and P3 were prepared as follows: P1 = 100 mM TRIS-HCl and 10 mM EDTA was prepared in 50 mL Milli-Q water and the pH adjusted to 8.0 with 1 M NaOH; P2 = 100 mM NaOH and 1% SDS was prepared to a volume of 50 mL in Milli-Q water; P3 = 3.5 M sodium acetate in Milli-Q water (50 mL). Cell samples from the expression time-course

were thawed and suspended in buffer P1 (250 μL). Buffer P2 (250 μL) was added to the cell solution and the mixture was gently agitated for 5 min. to dissolve the cells completely. The reaction was neutralized and SDS-debris precipitated by addition of buffer P3 (350 μL). The debris was collected by centrifugation at 14,000 rpm and the supernatent collected for analysis. A triplicate set of standard protein solutions of 0, 1, 2.5, 5, 7.5, 10, 12.5, and 15 μg/.1 mL alkaline aqueous BSA was prepared by dilution of the corresponding amounts of 1 μg/μL BSA stock to 100 μL with 1 M NaOH in a 96-well plate. Similarly, unknown protein samples (2.5 μL) were diluted in triplicate to 100 μL with NaOH. A working assay solution (100 μL) of 50:1 Bradford Reagent A:B (Thermo-Fisher Scientific) was added to each well, and the reaction was incubated for 2 hrs at 37°C prior to colorimetric absorbance analysis at 560 nm using a Spectromax 384 Plus UV/Vis Spectrophotometer (Molecular Devices).

Small-scale cell lysis and IMAC purification under denaturing conditions. Small scale lysis and purification of his-tagged proteins under denaturing conditions were performed according to the His-bind resin manufacturer (Novagen) protocols. The following buffers were prepared: A= 10 mM TRIS-HCl and 8 M urea prepared in 1 L Milli-Q water; B = buffer A supplemented with 100 mM phosphate at pH 8.0 in 50 mL Milli-Q water; C = buffer A supplemented with 100 mM phosphate at pH 6.3 in 50 mL Milli-Q water; D = buffer A supplemented with 100 mM phosphate at pH 5.9 in 50 mL

Milli-Q water; E = buffer A supplemented with 100 mM phosphate at pH 4.5 in 50 mL Milli-Q water. The cell paste from the end-point of the protein expression time-course was gently suspended in buffer B (8 mL/pellet from 500 mL growth) and gently rocked at 4 °C for 1 hr. The cellular debris was removed by centrifugation at 15,000-20,000 rpm for 45 min with a Sorvall SS-34 rotor, and the soluble fraction of the lysate was passed through a 1 mL bed of Ni(II)-charge IMAC resin. The column flow-through and subsequent washes with buffers C-E were collected separately and analyzed by SDS-PAGE/Coomassie staining.

Small-scale cell lysis and IMAC purification under native conditions. Small scale lysis and purification of his-tagged proteins under native conditions were performed according to the His-bind resin manufacturer (Novagen) protocols. Cell pellets were thawed and re-suspended in 2 mL lysis buffer (25 mM Tris•HCl, pH = 8.0, 1 mg/mL hen egg white lysozyme). The lysis was facilitated by rotary incubation at 4 °C for 1 hr, and the cellular debris was pelleted by centrifugation at 15,000 rpm for 20 min. The supernatant was loaded onto a Ni-charged IMAC resin as per the manufacturer's instructions. The column was washed with wash buffer (30 mM imidazole, 25 mM Tris•HCl, 0.01% Tween-20®, pH = 8.0), and the desired protein was eluted in elution buffer (500 mM imidazole, 25 mM Tris•HCl, pH =8.0). Samples of the column fractions

were subjected to SDS-PAGE and functional analysis of enzyme activity in the case of cellulase expression testing.

5.2 Evaluation of CBM3a•cellodextrin association

5.2.1 Cellodextrin synthesis and purification

Cellulose triacetate. Cellulose triacetate was prepared as previously reported with the following adapatations.[157] A solution of $HClO_4$ (70%, 2.2 mL), glacial acetic acid (540 mL), and acetic anhydride (60 mL) was stirred openly over an ice water bath. Oven-dried cellulose acetate (MW~30 kD, 39.4 wt% acetate content; 30 g) was added to the acid mixture and stirred openly at room temperature. After 2.5 hr, the reaction mixture was poured into 2.5 L ice-water and left overnight. The insoluble products were captured by vacuum filtration and washed with sat. $NaHCO_3$ until the filtrates were neutral by litmus paper. The crude reaction products were dissolved in CH_2Cl_2 (~600 mL) and washed with sat. $NaHCO_3$ (500 mL) and brine (300 mL). The organic layer was dried over anhydrous Na_2SO_4 and concentrated to afford product as a white solid (33 g). ^{1}HNMR ($CDCl_3$): δ 1.782 (bs, 1H), 1.945 (s, 3H), 2.010 (s, 3H), 2.130 (s, 3H), 3.548 (bs, 1H), 3.715 (t, 1H, *J* = 8.4), 4.063 (d, 1H, *J* = 5.6), 4.366-4.427 (m, 2H), 4.797 (t, 1H, *J* = 8), 5.071 (t, 1H, *J* = 9.2).

Pivalic anhydride. Pivalic acid (100 g, 0.98 mol) was dissolved in anhydrous triethylamine (200 mL, 1.47 mol) and diluted with anhydrous CH_2Cl_2 (200 mL). The

solution was stirred under Ar at 0°C. Trimethylacetyl chloride (120 mL, 0.98 mol) was added slowly to the pivalate solution. The reaction produced a white precipitate immediately and exothermally evolved HCl gas. The evolved gas was neutralized by bubbling through saturated ammonium hydroxide. The reaction mixture was diluted with CH_2Cl_2 (200 mL) and stirred at rt for 3hr. Anhydride formation was verified by TLC (10% EtOAc/ hexanes), and the reaction solids were removed by vacuum filtration. The solids were washed with ample amounts of Et_2O (1 L) and the filtrate was concentrated under reduced pressure. The resultant anhydride concentrate was purified by vacuum distillation to yield pure anhydride as a colorless liquid (185 mL, 99% yield). R_f = 0.4 (10% EtOAc: Hexanes). ^{1}HNMR ($CDCl_3$): δ 1.248 (s). ^{13}CNMR ($CDCl_3$): δ 26.57 (6C), 40.24 (2C), 173.95 (2C).

Pivaloylysis of cellulose triacetate. Pivaloylsis of cellulose triacetate was performed with slight modification of literature protocol.[157] Cellulose triacetate (5 g, 17 mmol) was dissolved in anhydrous CH_2Cl_2 (416 mL) at rt under an Ar. Pivalic anhydride (164 mL, 799 mmol) was added dropwise to the stirring solution over 30 min. BF_3•Et_2O (33.3 mL, 255 mmol) was added to the mixture and stirring continued at reflux under Ar. TLC (1:1 EtOAc: hexanes) and Maldi-MS monitoring showed ample reaction progress after 20 hr. The reaction was quenched with sat. $NaHCO_3$ until neutral by pH paper. The resultant solution was extracted into CH_2Cl_2 (3 X 330 mL), washed with brine (2 X 600 mL), and

dried over anhydrous Na_2SO_4. Concentration under slightly reduced pressure and then high vacuum yielded a crude, brown-green paint. This opaque material was diluted with minimal CH_2Cl_2 and the reaction products precipitated upon addition of hexanes. The solids were collected by vacuum filtration, combined with silica gel (10 g) in minimal CH_2Cl_2, and concentrated to dryness. The silicate mixture was purified on a by flash chromatography (220 g SiO_2, 75 ml/min; elution from 0 to 40% EtOAc: hexanes over 5 min, to 50% at 8 min, isocratic to 12 min, to 60% at 22 min, isocratic for 12 min, to 65% at 33 min, to 100% for 5 min) to afford reaction products as individual oligomers up to DP_6 and mixtures DP_7 and above.

Saponification of the pivaloylysate. The starting sugar (1 eq.) was dissolved in anhydrous methanol (0.1 mL/mg) under an argon atmosphere. 25% (wt/v) NaOMe/ MeOH (~0.3 eq.) was added to the sugar solution and the mixture stirred for 16 hours at room temperature. A white solid precipitated from the reaction mixture and partial saponification of the starting sugars was observed by LCMS. The reaction solids were collected by vacuum filtration and washed with cold MeOH. The intermediate products were dissolved in water (0.1 mL/mg), and 1.25 N LiOH (2 eq) was added. The reaction mixture stirred for 5 hours openly. The reaction was monitored by NP-TLC (4:2:1.5; EtOAc: MeOH: H_2O) and LCMS (NP-HPLC; 60% H_2O: MeOH; 50 °C; m/z = $(162n+18+Na)^+$). Upon complete saponification, the hydroxide was neutralized over

Dowex-[H^+] and the organic impurities extracted into ethyl acetate. The aqueous layer was freeze-dried to afford un-derivatized sugar.

Reversed-phase HPLC of protected cellodextrin mixtures. Protected cellodextrin oligomers were purified from the pivaloylysate of cellulose triacetate by RP-HPLC (C18-modified SiO_2, 65-68% ACN (0.2% Formic acid) in Water (0.2% formic acid); 42.5 mL/min over 35 min.) to yield clean, piv-capped products. m/z (ESI+) = 479 (M1+Na), 785 (M2+Na), 1073 (M3+Na), 1361 (M4+Na), 1649 (M5+Na), 1937 (M6+Na), 1124 (M7+2Na), 1266 (M8+2Na), 1410 (M9+2Na), 1554 (M10+2Na). m/z (MALDI+lin; matrix = HABA) = 2227 (M7+Na), 2515 (M8+Na), 2804 (M9+Na), 3092 (M10+Na), 3380 (M11+Na), 3668 (M12+Na). In addition to pivaloyl terminated cellodextrins, differentially terminated acetyl and pivaloyl capped cellodextrins up to DP_5 were isolated from RP-HPLC purification of the pivaloylysate. m/z (ESI+) = 1031 (M3+Na), 1319 (M4+Na), 1607 (M5+Na).

Mixed-acid hydrolysis of avicel. Avicel (20 g) was dissolved in concentrated HCl (37%, 160 mL) in an iced brine bath at -15 °C. Concentrated H_2SO_4 (>98%, 40 mL) was added slowly to the cellulose/acid slurry over 10 min. The slurry was stirred at rt for 3 hrs, or until the solution became homogeneous. The reaction solution was then poured into a reservoir of acetone (2 L) cooled to -20°C. The acetone slurry was left overnight, and the

cellulose precipitate was collected by filtration. The filter cake was dissolved in water (1 L). The acidic sugar solution was clarified by centrifugation at 5,000 rpm for 15 min with a Sorvall GS-3 rotor. The soluble sugar solution was filtered through basic Dowex in the [OH^-] form to remove excess chloride. The residual sulfate ions were removed via the adjustment of the acidic pH to 7.0 with $Ba(OH)_2$. The resulting $BaSO_4$ precipitate was removed by filtration over cotton and through a 0.22 μm nylon membrane. The clarified, neutral soluble sugar solution was freeze-dried to afford 5 g dry cellodextrins as a mixture.

Purification of un-derivatized cellodextrin mixtures. Dowex-50WX4 strong acid cation exchange resin (100 g) was generated in the Ca^{2+} form as previously described.[158] Biogel-P2 resin (300 g) was prepared in Milli-Q water in accordance with manufacturer instructions. A jacketed, glass chromatography column (1.2 m X 2.5cm ID) was fitted to a hot water bath (90 °C) and hot water was pumped through the column using a 3-MD-MT-HC model water pump (Little-Giant Pump Co., Oklahoma City, OK, USA). The degassed Biogel resin was poured into the column and allowed to settle while degassed water was flowed through the column openly. The degassed DOWEX resin was slowly poured directly on top of the Biogel resin in such a way as to not disturb the Biogel packing. Upon settling, the column was fit with a glass frit above the dowex resin. The resultant column was 90 cm in length (30 cm DOWEX and 60 cm Biogel). Underivatized

mixed-sugar solutions (<5 mL in hot Milli-Q water) were loaded directly onto the column via pipet transfer. Degassed water (~200 mL) was passed through the column by gravity flow (~1 – 2 mL/ min), and 1 mL fractions were collected, analyzed by NP-TLC (4:2:1.5; EtOAc: MeOH: H_2O), and like fractions pooled. The identity of single oligomers and/or multiple cellodextrins was confirmed by MALDI-MS (matrix = 2,5-dihydroxybenzoic acid; positive, linear mode): m/z = 527 (M3+Na), 689 (M4+Na), 851 (M5+Na), 1013 (M6+Na), 1175 (M7+Na), 1337 (M8+Na), 1499 (M9+Na), 1661 (M10+Na). Cellodextrin homogeneity was confirmed by LCMS (NP-HPLC; 75% H_2O: ACN; 50 °C).

5.2.2 BCA assay of cellodextrin concentration

Assay reagents A and B were prepared as described by Zhang and Lynd.[191] Standard cellobiose solutions were prepared in the concentration range of 2.5 - 75 μM and loaded in triplicate in a 96-well plate (100 uL sample/ well). Unknown cellodextrin solutions were prepared by dissolving ~50 - 100mg cellodextrin powder in buffer (5 – 10 mL) and preparing a 2.5:1000 dilution of this sample into a final volume of 100 μL. The freshly mixed working solution (100 μL; 1:1 solutions A:B) was added to the standard and unknown samples, and the mixtures were incubated at 72 °C for 35 min. The colored products were allowed to cool to room temperature and the absorbances at 560 nm were measured using a Spectromax 384 Plus UV/Vis Spectrophotometer (Molecular

Devices). The absorbance values were correlated to sugar concentrations using the Beer-Lambert law with a custom Microsoft Excel workbook.

5.2.3 Expression and purification of *C. cellulolyticum* CBM3a

CBM3a-encoding DNA was PCR amplified from a commercial vector (IDT, Coralvile, IA, USA) similar to pETscaf3 reported by Fierobe et al[287] using the forward primer TATA**CATATG**GCGGGTACCGGTGTTGTTTC with an NdeI cleavage site shown in bold and the reverse primer TATA**GTCGAC**TCACGGGGTAGAACCGTACG with a SalI restriction site in bold. CBM3a DNA was purified from an agarose gel (2 %) and subcloned into the pCR™-Blunt II-TOPO® (Invitrogen) vector for sequencing purposes. CBM3a DNA was isolated by restriction digestion and ligated into a pre-cut pET22b+ plasmid with T4 ligase forming pETCBM. pETCBM was transformed into chemically BL21(DE3) *E. coli* which were grown in LB media (12 L) supplemented with ampicillin (50 $mg \bullet L^{-1}$) at 37 °C to an OD_{600nm}= 0.4-0.8. The cells were then cooled to 30 °C and induced with IPTG (0.4 mM) for five hours and collected by centrifugation (5 krpm, 10 min). Cell pellets were typically stored at -78 °C for no more than 1 month prior to protein extraction. Cell pellets were lysed in lysis buffer (50 mM NaHEPES, 1 mM $CaCl_2$, 1 mM $SHCH_2CH_2OH$, 1 $\mu g \bullet mL^{-1}$ DNaseI, 1 mM PMSF, 20 μM PVS, pH = 7.5 - 7.9) by mechanical lysis. The cell debris was removed by ultracentrifugation at 40 krpm, 45 min, and 4 °C. The soluble fraction from the cleared crude lysate was passed over 20 g Avicel,

and CBM3a was elute with a linearly increasing gradient of pure water in 50 mM sodium phosphate buffer at pH = 7.5. The initially isolated protein was typically >90 pure by SDS-PAGE. Greater purity was achieved by elution of the protein concentrate (< 4 mL) through an S100HR gel filtration column in calorimetry buffer (50 mM NaHEPES, pH = 7.5, 1 mM $CaCl_2$, 1 mM β-mercaptoethanol) at a flow rate of 1.0 mL•min^{-1}. The protein was concentrated under N_2 pressure in a stirred-cell Amicon® (Amicon Inc., Beverly, MA, USA) using a 10 kMWCO PES membrane (EMD-Millipore) to a final concentration of 0.1 – 1 mM determined using the Edelhoch[286] method ($\varepsilon_{CBM,\ GndHCl}$ = 36,900 $cm^{-1}M^{-1}$).

5.2.4 Low-affinity ultrafiltration assay of CBM3a•cellodextrin association

The low-affinity CBM3a•cellopentaose and cellotetraose binding was assayed using the general procedure of Menguy et al.[168] Fresh CBM3a was concentrated in calorimetry buffer by centrifugal concentration through a Corning 10 kMWCO PES membrane (Thermo-Fisher Scientific) to a final concentration of 280 μM. CBM3a and ligand stocks solutions were used to prepare a range of 0.5 mL mixtures from 0.025 mM to 24 mM cellodextrin with 0.028 mM CBM3a. A set of control samples lacking the CBM was also prepared. The samples were incubated over night at 4 °C and then concentrated by centrifugation at 15,000 g in 0.5 mL PES concentrators until the protein mixture had concentrated to about 25 μL (20X). The free ligand concentration in the flow

through was measured using the BCA assay, and the obtained value was used to determine the amount of ligand bound/ pooled on the protein side of the membrane via the following relationships.

$$K_d = \frac{[P_f][L_f]}{[PL]}$$

Equation 43

$$[P_{Total}] = [P_f] + [PL]$$

Equation 44

$$[L_{Total}] = [L_f] + [PL]$$

Equation 45

$$\theta = \frac{[PL]}{[P_{Total}]} = \frac{[L_f]}{K_d + [L_f]}$$

Equation 46

The fraction of bound protein (θ) was plotted against the free ligand concentration in the Origin 5 platform (GE Healthcare) to determine the apparent K_d via Equation 46.

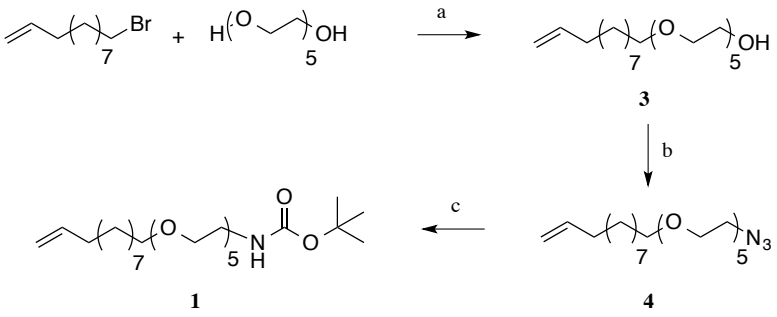
5.3 Force spectroscopy of MiniCipC-cellulose interactions

5.3.1 Cellulose surface preparation

Cellulose nanocrystal suspensions were prepared and spin-cast onto a polymer pre-coated silicon wafer as reported by Wågberg et al.[183] AFM images were collected in

tapping mode using bare Ultrasharp NSC 15/AIBS tips (Mikromasch, Lady's Island, SC, USA).

5.3.2 Synthesis of heterobifunctional linker molecules 2.1 and 2.2



a) pentaethylene glycol, NaH, THF, 0 °C – 65 °C, 2 hr, then alkene, reflux, 12 hr; b) DBU, DPPA, DMF, 0 °C – rt, 1.5 hr, then NaN_3, 90 °C, 5 hr; c) Ph_3P, THF, H_2O, 15 hr, then Boc_2O, Et_3N , 0 °C—rt, 4 hr.

Scheme 8: Synthesis of *tert*-Butyl 3,6,9,12,15-pentaoxahexacos-25-enylcarbamate (2.1)

5.3.2.1 *tert*-Butyl 3,6,9,12,15-pentaoxahexacos-25-enylcarbamate (2.1).[179]

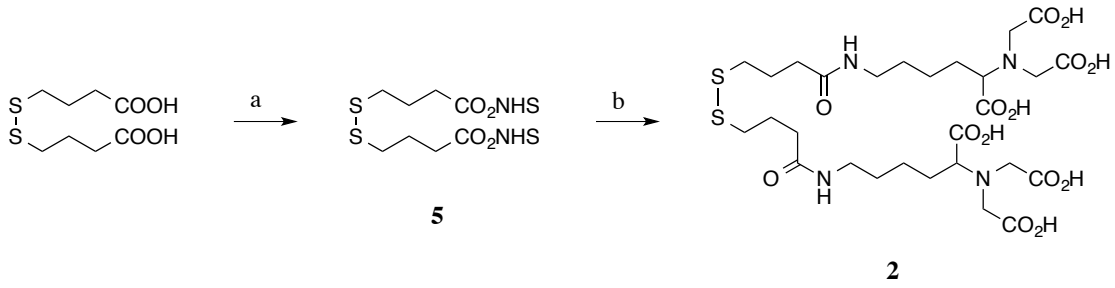
The title compound was synthesized as outlined in Scheme 8 via the method of Shestopalov[179] with minor modifications. Pentaethylene glycol (20.1 g, 84.3 mmol) was dissolved in dry THF (100 mL) and stirred on ice under N_2 atmosphere. NaH (1.01 g, 60% dispersion in mineral oil, 25.3 mmol) was added to PEG solution. The mixture warmed to room temperature over 30 min and was then refluxed for 2 hrs, at which point it turned orange. A solution of 11-bromo-1-undecene (5 mL, 23 mmol) in THF (20 mL) was added slowly to the reaction over 30 min while hot, and the reaction continued to stir at reflux over night. After 12 hrs, the solvent was removed under reduced

pressure, and the crude residue was triturated in hexanes (5 X 150 mL). The extracts were combined and washed with brine (2 X 200 mL), dried over Na_2SO_4 and concentrated to a pale yellow oil (**2.3**, 7.03 g, 78% yield) that was used in the subsequent step without further purification. ESI(+)-MS: $[M+H]^+$, expected m/z = 391.56; found m/z = 391.

Crude **2.3** (8.85 g, 22.7 mmol) was dissolved in dry DMF (20 mL) and cooled to 0 °C under Ar. 1,8-diazabicyclo[5.4.0]undec-7-ene (6.8 mL, 45.4 mmol) and diphenyl phosphorazidate (9.8 mL, 45.4 mmol) were added and the mixture stirred on ice for 1.5 hrs. Sodium azide (3.7 g, 56.8 mmol) was added and the reaction was warmed to 90 °C for 5 hrs. The DMF was removed under reduced pressure, and the residue was combined with water (50 mL) and extracted with Et_2O (300 mL). The extracts were washed with water (2 X 50 mL), sat. NH_4Cl (50 mL), and brine (2 X 50 mL). The organic layer was dried over Na_2SO_4 and concentrated to a brown oil (13 g). The crude was dry loaded onto SiO_2 (15 g) and purified by flash chromatography over SiO_2 (220 g, 40 mL/min) by eluting 0 – 65% EtoAc: hexanes to remove impurities, then elution to 80% EtOAc: hexanes to elute desired product **2.4 (**2.73 g, 6.6 mmol, 29% yield) and the intermediate phosphinate ester (2.64 g, 4.25 mmol, 19% yield; m/z = 623 $[M+H]^+$) which could be recycled to the desired product (1.5 g, 3.61 mmol, 85% recovery, 48% combined yield) by heating with sodium azide (690 mg, 10.6 mmol) in DMF (10 mL) at 90 °C until

complete by TLC (1:1 EtOAc: hexanes; PMA stain). ESI(+)-MS: $[M+Na]^+$, expected m/z = 438; found: 438.

2.4 (2.73 g, 6.6 mmol) was dissolved in THF (20 mL), and triphenylphosphine (1.91 g, 7.3 mmol) and water (178 μL, 9.9 mmol) were added. After stirring for 12 hrs, TLC (1:1 EtOAc: hexanes) indicated remaining starting material. The reaction was warmed to reflux for 2 hrs, and the solution was cooled to 0 °C. Triethylamine (1.08 mL, 7.9 mmol) was added, and mixture became milky white. Di-*tert*-butyl dicarbonate (1.82 mL at 37 °C, 7.9 mmol) was added drop-wise while warm to the reaction mixture. The reaction warmed to rt and stirred for 4 hrs. Petroleum ether (200 mL) was added, and the triphenylphosphine oxide precipitate was filtered away. The solvents were removed from the filtrate under reduced pressure, and the crude residue was purified over SiO_2 (50 g) packed in pet ether. The title compound eluted at 20% acetone in petroleum ether as a colorless oil (**2.1**, 2.97 g, 6.07 mmol, 92% yield). ^{1}H NMR (400 MHz, cdcl$_3$) δ 5.81 (apparent ddt, *J* = 16.9, 10.2, 6.7 Hz, 1H), 5.10 (bs, 1H), 5.03 – 4.89 (m, 2H), 3.70 – 3.51 (m, 18H), 3.44 (t, *J* = 6.8 Hz, 2H), 3.35 – 3.28 (m, 2H), 2.08 – 2.00 (m, 2H), 1.61 – 1.53 (m, 2H), 1.44 (s, 9H), 1.38 – 1.22 (m, 12H). ^{13}C NMR (75 MHz, cdcl$_3$) δ 155.93, 139.09, 114.07, 78.99, 71.45, 70.56, 70.55, 70.54, 70.47, 70.20, 70.16, 70.00, 40.32, 33.74, 29.58, 29.47, 29.40, 29.36, 29.05, 28.86, 28.38, 26.03. ESI(+)-MS: $[M+Na]^+$, expected m/z = 512; found: 512.5.



a) NHS (2 eq), DCC, dioxane, 18 hr, rt; b) NTA-Lysine (2 eq), $NaHCO_3$, H_2O, DMF, pH = 7.2, 12 hr, rt.

Scheme 9: synthesis of 2,25-bis(carboxymethyl)-9,18-dioxo-13,14-dithia-2,8,19,25-tetraazahexacosane-1,3,24,26-tetracarboxylic acid (2.2).

5.3.2.2 2,25-bis(carboxymethyl)-9,18-dioxo-13,14-dithia-2,8,19,25-tetraazahexacosane-1,3,24,26-tetracarboxylic acid (2.2).

2.5 was prepared as reported by Ludwig and Jay.[288] **2.5** (500 mg, 1.15 mmol) in DMF (30 mL) was added to a stirring solution of NTA-Lysine (Hydrate, 949 mg, 3.5 mmol; Sigma-Aldrich, St. Louis, MO, USA) in buffer (15 mL; 20 mM $NaHCO_3$, pH = 7.2) for 12 hrs, whereby the reaction was complete by analytical LCMS (3 – 43% MeCN: H_2O + 0.02% formic acid over 15 min at 1.0 mL/ min; 4.6 X 250 mm C18-silica; Agilent Technologies, Santa Clara, CA, USA). The reaction was concentrated under vacuum and purified by RP-HPLC over 22 x 250 mm C18 column using a 26.2 mL/min flow rate and the same elution as in analytical LCMS. Product fractions were pooled and concentrated under vacuum to remove MeCN and acid. The crude was dissolved in water and freeze-dried to a white, fluffy powder (336 mg, 0.46 mmol, 40% yield). ^1H NMR (400 MHz, d_2o)

δ 4.10 – 3.92 (m, 8H), 3.25 – 3.09 (m, 4H), 2.97 (s, 1H), 2.81 (s, 1H), 2.68 (bt, *J* = 6.8 Hz, 4H), 2.31 (bt, *J* = 6.6 Hz, 4H), 2.01 – 1.79 (m, 8H), 1.61 – 1.40 (m, 8H). ^{13}C NMR (75 MHz, d_2o) δ 175.74, 172.35, 170.19, 67.98, 55.19, 38.77, 37.00, 34.24, 28.00, 26.46, 24.64, 23.56. ESI(-)-MS: $[M-H]^-$, expected m/z = 725.81; found: 725.6.

5.3.3 2.2 disulfide reduction monitored by UV/Vis spectroscopy and Ellman's test for sulfhydryl reactivity.

Our initial attempts to observe thiol ligation into our extended surfaces by XPS were unsuccessful; therefore, we first sought to confirm that the sulfhydryl group was formed upon reaction with TCEP. As depicted in Figure 61, the S2 disulfide reduction in the presence of equimolar TCEP in alkaline buffer results in an immediate a loss of UV absorbance at 274 nm.

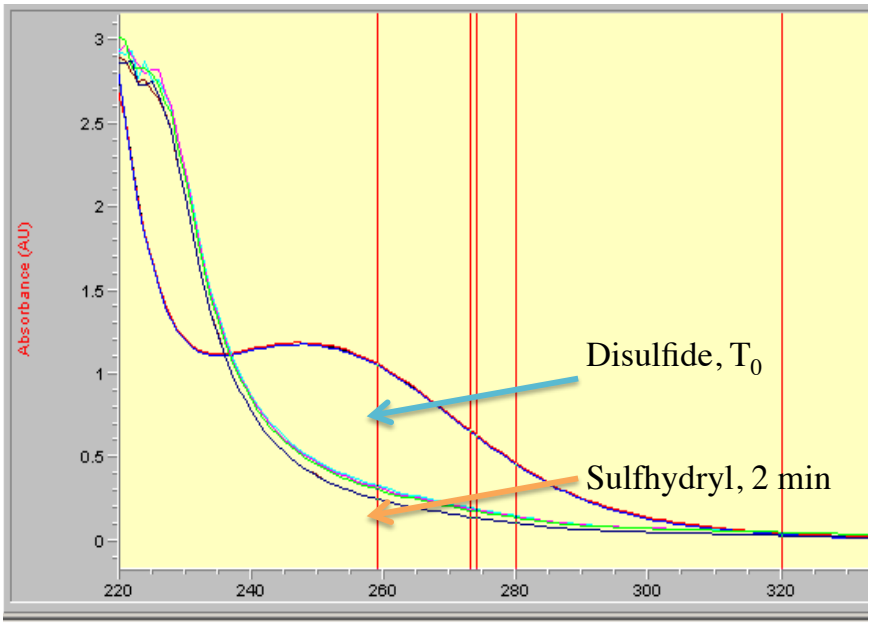


Figure 61: 2.2 disulfide reduction with TCEP monitored by UV/Vis spectroscopy

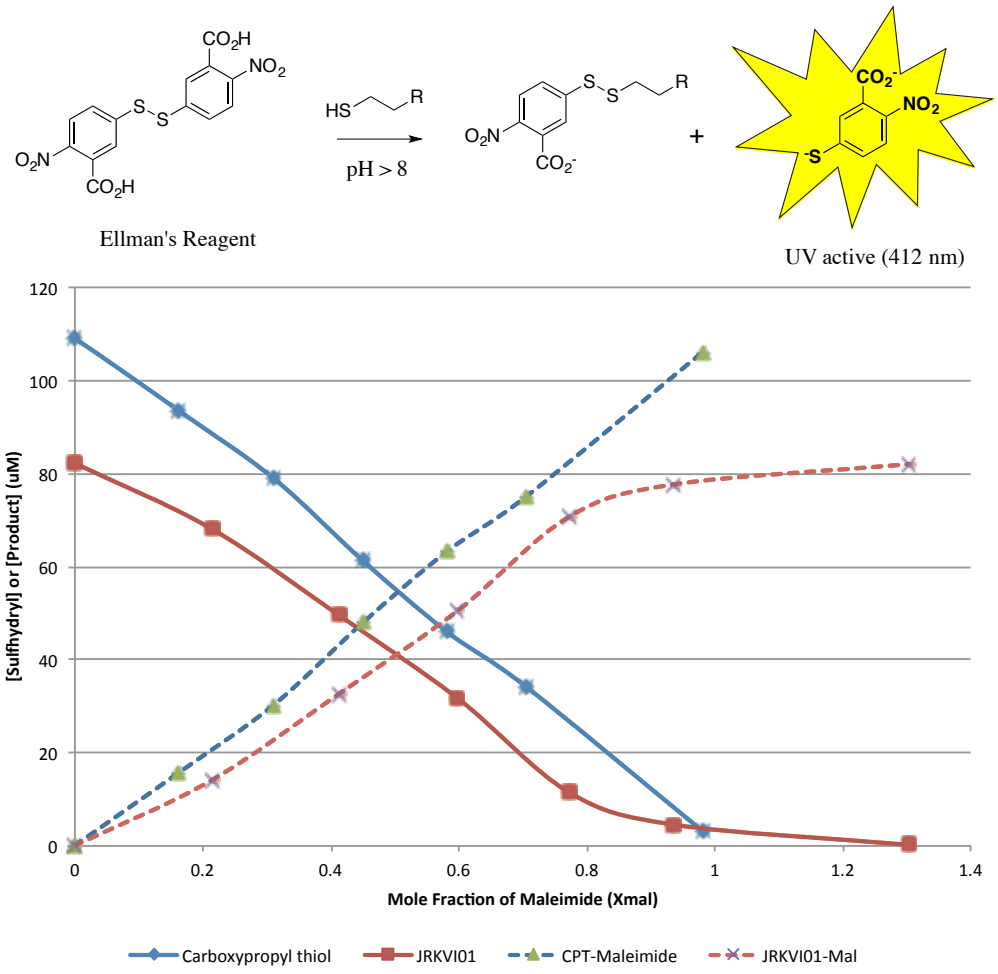


Figure 62: Ellman's test of sulfhydryl reactivity for 2.2-thiol

Indeed, the appearance of the sulfhydryl group does not necessarily prove reactivity with a maleimide cross-linker. We next used performed an Ellman's test (Thermo Scientific of Pierce Biotechnology, Rockford, IL, USA) of free sulfhydryl groups in the presence of increasing concentration of 4-maleimidobutryate (Mal). A solution of **2.2** (JRKVI01, 10 mM) and equimolar TCEP (pH = 8.0) were assayed for free thiol content with varying amounts of Mal up to 1.3 equivalents. To ensure the reactivity of the

Ellman's reagent with free thiols, a solution of 3-carboxypropyl disulfide (5 mM) and equimolar TCEP was prepared as a positive test for reactive thiols. The results are shown in Figure 62, confirming a 1:1 reaction between the reduced **2.2** and maleimide.

5.3.4 Solid Phase Peptide Synthesis of a hexahistidine peptide

Solid phase peptide synthesis was carried out on a Liberty model microwave assisted peptide synthesizer (CEM, Mathews, NC, USA). Reagents were prepared for a 0.1 mmol scale synthesis of CGWGGHHHHHH. Fmoc-protected amino acids and Rink resin were purchased from NovaBiochem® (EMD-Millipore, Billerica, MA, USA). Sequential addition of Fmoc-C(Trt)-OH (590 mg in 5 mL DMF), Fmoc-G-OH (480 mg in 8 mL DMF), Fmoc-W(Boc)-OH (530 mg in 5 mL DMF), and Fmoc-H(Trt)-OH (1.98 g in 16 mL of 20% DMSO:DMF) to the Rink resin (100-200 mesh; 170 mg) was performed via the automated controls of the synthesizer. HOBt (450 mM in 13 mL DMF) and DIEA (2 M in 4.56 mL NMP) were used for the activator and activator base, respectively. The Fmoc deprotection solution was 20% (v/v) piperidine with 0.1 M HOBt in DMF (207 mL). Upon completion of the coupling reactions, the peptide was cleaved from the resin manually with 5 mL 94:1:2.5:2.5 [trifluoroacetic acid: triisopropylsilane: ethanedithiol: H_2O] per 65 mg resin. After shaking at room temperature for 3 hrs in a glass scintillation vial, the bright yellow reaction mixture was filtered under vacuum and the resin washed with methanol. The filtrate was concentrated to an oil under vacuum and poured over ethyl ether (60 mL at -20 °C) to precipitate the peptide. The precipitate was dissolved in

H_2O and lyophilized to a white powder (11.7 mg; yield is with respect to ½ of reaction mixture). The peptide mass was confirmed by MALDI-TOF analysis. Matrix = HCCA in 1% TFA solution; $[M+H]^+$, predicted: 1301.52, found: 1300.83; $[M+Na]^+$, predicted: 1323.51, found: 1322.53. The peptide was used in AFM without further purification.

5.3.5 Expression and purification of His_6-tagged miniCipC

MiniCipC-encoding DNA was PCR amplified from a commercial vector (IDT, Coralvile, IA, USA) similar to pETscaf3 reported by Fierobe et al[287] using the forward primer CATA**CATATG**GCGGGTACCGGTGTTGTTTCTGTTC with an NdeI cleavage site shown in bold and the reverse primer TATA**GTCGAC**AACCGCAACTTTCAGTT-CTTTGGTCG with a SalI restriction site in bold. MCC DNA was purified from an agarose gel (0.8%) and subcloned into the pCR™-Blunt II-TOPO® (Invitrogen) vector for sequencing purposes. MCC DNA was isolated by restriction digestion and ligated into a pre-cut pET22b+ plasmid with T4 ligase forming pETMiniCipC. pETMiniCipC was transformed into electrocompetent BL21(DE3) *E. coli* which were grown in LB media supplemented with ampicillin (50 $mg \bullet L^{-1}$) at 37 °C to an OD_{600nm}= 0.4-0.8. The cells were then cooled to 30 °C and induced with IPTG (0.4 mM) for seven hours and collected by centrifugation (5 krpm, 10min). Cell pellets were typically stored at -78 °C for no more than 1 month prior to protein extraction. Cell pellets were lysed in lysis buffer (50 mM NaHEPES, 1 mM $CaCl_2$, 1 mM $SHCH_2CH_2OH$, 1 $\mu g \bullet mL^{-1}$ DNaseI, 1 mM PMSF, 20 μM PVS, pH = 7.5 - 7.9) by mechanical lysis. The cell debris was removed by

ultracentrifugation at 40 krpm, 45 min, and 4 °C. The soluble fraction from the cleared crude lysate was passed over 20 g Avicel, and MCC was elute with a linearly increasing gradient of pure water in 50 mM sodium phosphate buffer at pH = 7.5. The initially isolated protein was typically >90 pure by SDS-PAGE. Greater purity was achieved by elution of the protein concentrate through an S100HR gel filtration column in phosphate buffered saline (pH = 7.5) with 0.05% Tween-20® at a flow rate of 1.0 mL•min^{-1}. The protein was concentrated under N_2 pressure in a stirred-cell Amicon® (Amicon Inc., Beverly, MA, USA) using a 10 kMWCO PES membrane (EMD-Millipore) to a final concentration of ~40 μM determined using the Edelhoch[286] method ($\varepsilon_{MCC, GndHCl}$ = 42,860 cm^{-1}M^{-1}).

5.3.6 Expression and purification of His_6-tagged murine galectin-3

Murine galectin-3 was expressed and purified as described by Bowers et al[184] with minor modifications. pETG3 was expressed in BL21(DE3) *E. coli* with 1 mM IPTG at 37 °C for 3.5 hrs, and the cells were collected by centrifugation at 5 krpm for 10 min at 4 °C. The cells were suspended in a hand-held glass tissue grinder with lysis buffer (50 mM Na Phosphate, 1 μg•mL^{-1} DNaseI, 1 mM PMSF, pH = 7.5) and lysed mechanically with an EmulsiFlex-C5 homogenizer. Cellular debris was removed at 20,000 x g for 30 min at 4 °C, and the supernatant was passed through a 5 mL bed of Ni-loaded NTA His-bind® resin (EMD-Millipore) as per the manufacturer's instructions. Pure G3 eluted from the column with 100 mM imidazole. The fractions appearing pure by SDS-PAGE were

pooled and concentrated to ~10 μM using 3 kMWCO centricon spin-concentrator (EMD-Millipore). The remaining imidazole was removed by dialysis (3 X 8 hrs) against phosphate buffered saline with Tween-20® through a 3.5 kMWCO snakeskin dialysis tubing (Thermo Scientific, Rockford, IL, USA). Some precipitate formed during dialysis. The final protein solution was clarified by centrifugation (4 krpm, 10 min, 4 °C) and concentrated to ~40 μM ($\varepsilon_{G3,\ GndHCl}$ = 29,800 $cm^{-1}M^{-1}$).

5.3.7 AFM cantilever functionalization

The AFM cantilevers were functionalized as described via a combined approach of Bowers[180] and Shestopalov.[179] In a clean environment, AFM tips (NP, nominal spring constant of 0.120 N/nm; Bruker, Billerica, MA) were soaked in Nano-Strip® at 75 °C for 30 s, rinsed with ultrapure water, and dipped into 5% aqueous HF for 1 min. The acid was blown away with N_2, and the tip immediately coated with 10 μL alkene **2.1**. The tip(s) were transferred to a nitrogen-filled glove box and irradiated with UV light for 30 min at room temperature (UVP 11sc lamp, 4400 $\mu C \bullet cm^{-2}$ at 5 cm distance above tips). The tips were then removed from the cleanroom, rinsed with filtered EtOH, Miili-Q H_2O, EtOH, and CH_2Cl_2. The tips were submerged in 50% trifluoroacetic acid in CH_2Cl_2 for 30 min then rinsed with EtOH, Milli-Q H_2O, EtOH and CH_2Cl_2. The tips were submerged in a solution of NHS-$dPEG_{24}$®-Mal (15 mM, 1% Et_3N in CH_2Cl_2; Quanta BioDesign, Powell, OH, USA) for 2.5 hrs and rinsed with CH_2Cl_2 and EtOH. The tips were transferred to a freshly prepared solution of NTA-thiol **2.2** (10 mM with respect to

reduced sulfhyryl, 5 mM TCEP, 10 mM aqueous Na Phosphate, pH = 7.5, filtered 0.2 μm) for 13 – 16 hrs. The tips were rinsed with Milli-Q H_2O and soaked into nickel solution (50 mM $NiSO_4$, 20 mM NaHEPES, pH = 7.0) for 30 min and rinsed with water. The tips were soaked in bind buffer (5 mM imidazole, 10 mM NaPhosphate, 0.15 M NaCl, 0.05% Tween-20®, pH = 7.5) for 15 – 30 min and then transferred to the protein solution (~250 μL per tip; ~40 μM protein in PBS-T (10 mM NaPhosphate, 0.14 M NaCl, 0.05% Tween-20®, pH = 7.4) using a mass spectrometry vial cap placed in a covered glass weigh dish as the reaction vessel. The protein soaking step was omitted in the blank tip and His_6 unbinding experiments. The tips were rinsed with phosphate-buffered saline immediately before the un-binding experiments.

5.3.8 Functionalization of silicon surfaces with His_6 peptide

In a cleanroom, 1 cm^2 silicon wafers were soaked in NanoStrip® for 10 min at 75 °C, rinsed with water, and soaked in 5% aqueous HF for 5 min at rom temperature. The acid was blown away under a N_2 stream, and the surface immediately covered with alkene **2.1**. The surface was transferred to a N_2-filled glove box and irradiated with UV light for 2 hrs at room temperature (UVP 11sc lamp, 4400 $\mu C \bullet cm^{-2}$ at 2 cm distance above surface). The surface was removed from the glove box and clean room and rinsed with CH_2Cl_2, EtOH, water, EtOH, and CH_2Cl_2. The surface was dried under Ar and soaked in 50% TFA:CH_2Cl_2 for 30 min. The acid was rinsed away with water, EtOH, CH_2Cl_2 and the surface was soaked in the NHS-$dPEG_{24}$®-Mal solution as described above

for 2.5 hrs. The surface was rinsed with CH_2Cl_2, EtOH, and water and placed in a 0.2 μm- filtered solution of the CGWGGHHHHHH peptide (4 mM in 20% DMSO:water, 10 mM TCEP, pH = 7.5; Supporting Information) for 16 hrs. The surface was loaded with NaHEPES-buffered $NiSO_4$ as described above prior to use in His_6 un-binding experiments.

5.3.9 Goniometric pH titration of surface 2.6

A 1 cm^2 silicon wafer was functionalized with an NTA modified monolayer as described above. A series of 10 mM buffers were used to generate monotonic increases in pH from 2.03 – 11.05: phosphoric acid/ monobasic sodium phosphate (pH = 2 – 3); acetic acid/ sodium acetate (pH = 4 – 5); monobasic/ dibasic sodium phosphate (pH = 6 – 8); sodium bicarbonate/ carbonate (pH = 9 – 11). Advancing and receding contact angles for a drop of buffer on the NTA surface were collected in triplicate using an NRA C. A. model 100 contact angle goniometer. The average contact angles are given in Table 5 and shown here in Figure 63. A clear decrease in contact angle with increasing pH is consistent with greater surface wetting as the NTA moiety is ionized. Although the reported pK_a of NTA is ~3 (pK_a = 3.03, 3.07, 10), the surface appears to perturb the pK_a to ~5.5. As a result, buffered $NiSO_4$ at pH > 6.5 is necessary to efficiently immobilize Ni(II) to the surface.

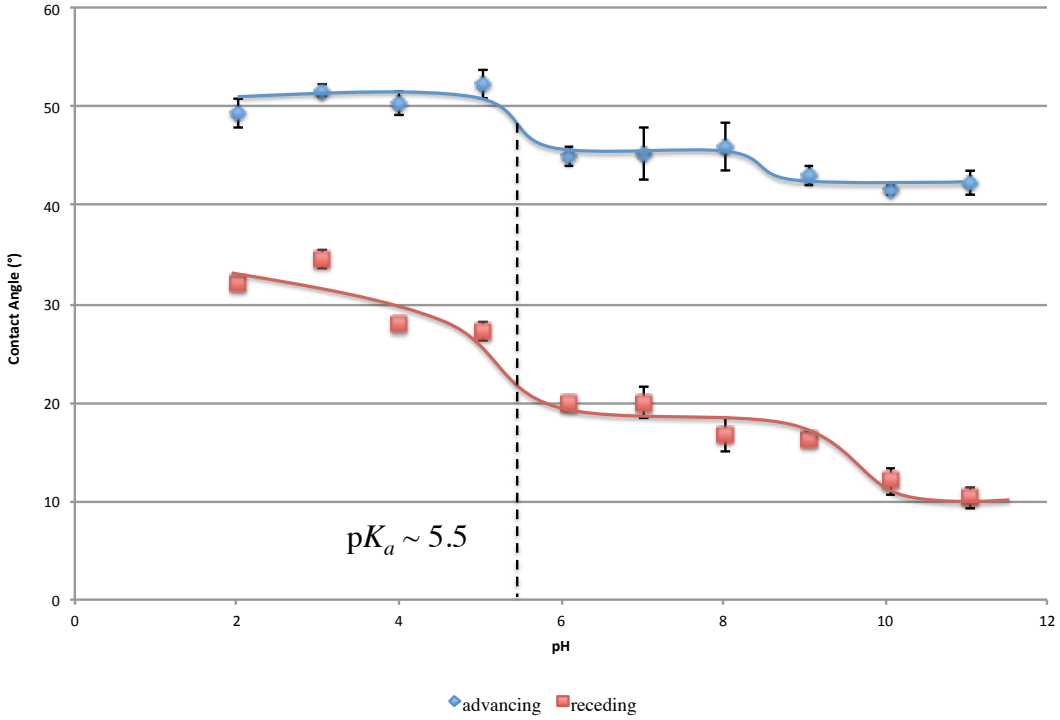


Figure 63: Goniometric pH titration of surface 2.6

5.3.10 Unbinding experiments

The unbinding experiments were performed as outlined by Bowers et al.[184] Automated pulling experiments were carried out on a custom-built 3-axis AFM composed of a MultiMode head (Digital Instruments, Santa Barbara, CA, USA) mounted on an xy- and z-positioning stages (Physik Instrumente, Auburn, MA, USA).[289, 290] At least 250 pulls were generated and the data collected using an automated program.[289, 290] The sample surface was brought into contact with the tip, held for a dwell time of 1 s, and then retracted at a constant velocity of 200 nm•s^{-1}. The force loading rate (383 ± 183

$pN \bullet s^{-1}$) was determined from the cantilever spring constant (120 $pN \bullet nm^{-1}$) provided by the manufacturer and the slope on the force curve before an unbinding event; see Supporting Information of Bowers et al.[184] As per our prior findings, pulls with contact forces in excess of 350 pN or below 30 pN were removed for analysis. Unblocked protein experiments (initial and washed) and His_6 un-binding experiments were carried out in the presence of PBS-T (30 μL, see AFM cantilever functionalization procedure). For the blocked experiments, the flow cell was initially flooded with 30 μL cellulose NC suspension (1.6 wt% in PBS-T, filtered through 0.2 μm PES membrane). After 10 min, the NCs were removed with PBS-T (10 X 30 μL), and the data collection continued. Then, the NCs were removed with water (10 X 30 μL) and PBS-T (3 X 30 μL) and data collection continued for the washed experiments. The photodetector was calibrated prior to data collection so as not to foul the tip with excessive force. The data was analyzed using a custom Matlab (MathWorks, Natick, MA, USA) script[290] and the resulting rupture force and length histograms were fit to inverse Gaussian* distributions using the Matlab curve fitting toolkit (Appendix A).

* Though the inverse Gaussian function more qualitatively matched the generated histograms, the means obtained with Normal distribution functions were nearly identical in each case to the means from the Gaussian function.

5.4 Production of recombinant cellulases and reducing sugar assays

5.4.1 Expression and purification of wild type and chimeric C. cellulolyticum endoglucanase Cel9G

pET9Gc and pET9Gt plasmids were generously gifted from Ed Bayer and Henri-Pierre Fierobe. The plasmids were transformed into BL21(DE3) *E. coli* via electroporation, and a single colony was used to inoculate a starter culture of 5 mL LB broth supplemented with 50 mg•L^{-1} ampicillin. The culture was grown overnight at 37 °C and then added to 50 mL LB/Amp broth. The culture was agitated at 200 rpm and 37 °C until the cell density increased to $OD_{600\ nm}$= 1.0. A 20 mL aliquot of the starter culture was added to ampicillin-containing auto-induction media (500 mL) and the cells were agitated at 200 rpm and 25 °C for 23 hrs. The cells were collected by centrifugation (5,000 rpm, 10 min, 4 °C) and stored at -78 °C. The cell pellets were thawed on ice and re-suspended in lysis buffer (25 mM Tris•HCl, pH = 8.0, 1 mg•mL^{-1} DNaseI, 1 mM PMSF, 20 μM PVS) with a hand-held tissue grinder. The cells were disrupted by mechanical lysis in an EmulsiflexC5 homogenizer and then held on ice for 45 min. The cell debris was pelleted by ultracentrifugation (40,000 rpm, 45 min, 4 °C) and the supernatant was applied to a Ni-charged His-bind IMAC column. The protein was isolated by FPLC. A linear gradient from 5 – 30 mM imidazole in Tris buffer (pH = 8) over 100 mL at 4 mL•min^{-1} was applied to remove cellular impurities, and the continued elution to 150 mM imidazole was required to elute His_6-containing Cel9G. The protein was impure by

SDS-PAGE, and the Cel9G fractions were pooled, concentrated, and buffer exchanged into Q-Sepharose bind buffer (25 mM Tris•HCl, pH = 8.0, 5 mM β-mercaptoethanol, 1 mM EDTA) using a 30 kMWCO PES membrane centrifugal concentrator (Thermo-Fisher Scientific). During the third exchange, the solution was concentrated to 5 mL and loaded onto a pre-equilibrated Q-Sepharose FPLC column (2.5 X 15 cm). The protein was isolated in pure form (as analyzed by SDS-PAGE with Coomassie stain) from the column upon linear elution from 0 – 230 mM NaCl in bind buffer at 5 $mL \bullet min^{-1}$. The pure fractions were concentrated to ~2 mL, supplemented with Pepstatin A (1 $\mu g \bullet mL^{-1}$), and stored at 4 °C. Protein concentration of the concentrate was 10 – 20 μM as determined by the Edelhoch method assuming $\varepsilon_{CelG,\ GndHCl} = 176{,}000\ cm^{-1}M^{-1}$.

5.4.2 Expression and purification of wild type and chimeric C. cellulolyticum cellobiohydrolase Cel48F

pET48Fc and pET48Ft plasmids were generously gifted from Ed Bayer and Henri-Pierre Fierobe. The plasmids were transformed into BL21(DE3) *E. coli* via electroporation, and a single colony was used to inoculate a starter culture of 5 mL LB broth supplemented with 50 $mg \bullet L^{-1}$ ampicillin. The culture was grown overnight at 37 °C and then added to 1 L LB/Amp broth. The culture was agitated at 200 rpm and 37 °C until the cell density increased to $OD_{600\ nm} = 1.9$. The culture was cooled to 17 °C and induced with IPTG at a final concentration of 40 μM. The cells were agitated at 200 rpm and 17 °C for 16 hrs. The cells were collected by centrifugation (5,000 rpm, 10 min, 4 °C)

and stored at -78 °C. The cell pellets were thawed on ice and re-suspended in lysis buffer (25 mM Tris•HCl, pH = 8.0, 1 mg•mL^{-1} DNaseI, 1 mM PMSF, 20 μM PVS, 1 μg•mL^{-1} Pepstatin A) with a hand-held tissue grinder. The cells were disrupted by mechanical lysis in an EmulsiflexC5 homogenizer and then held on ice for 45 min. The cell debris was pelleted by ultracentrifugation (40,000 rpm, 45 min, 4 °C) and the supernatant was applied to a Ni-charged His-bind IMAC column. The protein was isolated by FPLC. A linear gradient from 5 – 30 mM imidazole in Tris buffer (pH = 8) over 100 mL at 4 mL•min^{-1} was applied to remove cellular impurities, and the continued elution to 150 mM imidazole was required to elute His_6-containing Cel48F. The protein was impure by SDS-PAGE, and the Cel48F fractions were pooled and concentrated to 800 μL using a 30 kMWCO PES membrane centrifugal concentrator (Thermo-Fisher Scientific). The concentration was loaded onto a HR-S100 FPLC gel filtration column (2.5 X 60 cm) that was pre-equilibrated with Cel48F gel filtration buffer (25 mM Tris•HCl, pH = 8.0, 100 mM NaCl, 5 mM b-mercaptoethanol, 1 mM EDTA, 1 mM PMSF, and 1 μg•mL^{-1} Pepstatin A). The protein was isolated in pure form (as analyzed by SDS-PAGE with Coomassie stain) from the column between 100 – 130 mL of eluent at 1.2 mL•min^{-1}. The pure fractions were concentrated to ~2 mL, buffer exchanged into Cel48F storage buffer (25 mM Tris•HCl, pH = 8.0, 100 mM NaCl, and 1 μg•mL^{-1} Pepstatin A), and stored at 4 °C. Protein concentration of the concentrate was 20 – 40 μM as determined by the Edelhoch method assuming $\varepsilon_{CelF,\ GndHCl}$ = 211,665 cm^{-1}M^{-1}.

5.4.3 Reducing sugar assays of cellulase activity

In general, cellulase activity on insoluble avicel (aka avicelase acvitity) was determined by the discontinuous colorimetric quantitation of reducing sugar content from the soluble fraction at predetermined reaction time points. Cellulase samples (200 μL) were combined with 1.8 mL of avicelase reaction mixture (8 g•L^{-1} avicel, 20 mM Tris Maleate, pH = 6.04, 100 mM NaCl, 10 mM $CaCl_2$), and the slurry was incubated while shaking (200 rpm) at 37 °C. Aliquots (0.4 mL) were removed after 1, 3, 6, and 16 hrs, and the reaction was quenched by submersion in a boiling water bath for 5 min. The insoluble fraction was pelleted by centrifugation at 13,000 rpm for 10 min and the supernatant was assayed for reducing sugar content using one of the methods detailed below and protein content using the BCA protein quantitation protocol with BSA as the standard (Thermo-Fisher Scientific).

Ferricyanide or the "Prussian Blue" method. The reducing sugar assay was performed as described by Park and Johnson.[190] The avicelase supernatant (400 μL) was combined with 400 μL ferricyanide solution (0.05 wt% aqueous potassium ferricyanide) and 400 μL carbonate-cyanide solution (1.33 g Na_2CO_3 + 0.163 g KCN in 250 mL water). The mixture was boiled in water for 15 min and cooled to rt. The mixture was combined with 2 mL ferric iron solution (0.015 wt% ferric ammonium sulfate and 0.01% SDS in 0.025 M H_2SO_4) for 15 min at rt. Enzyme cross-reaction led to blue precipitate formation that was removed by centrifugation at 13,000 rpm for 5 min. The blue color in the supernatant

was quantified spectrophotometrically at 700 nm in comparison to a glucose standard curve.

Dinitro salicylic acid method. The reducing sugar assay of Miller[291] was used as described by Moretti and Thorson.[189] The avicelase supernatant (250 μL) was added to 750 μL of DNS reagent (1 wt% DNS, 0.2 wt% phenol, and 0.05 wt% $NaHSO_3$ in 1 wt% aqeous NaOH). The mixture was heated in a glass tube at 100 °C for 5 min, and upon cooling, the absorbance at 525 nm was measured against a glucose standard curve.

BCA reducing sugar assay. The BCA reducing sugar assay was performed as described above for the quantitation of cellodextrin concentrations in CBM3a•cellodextrin ITC studies.

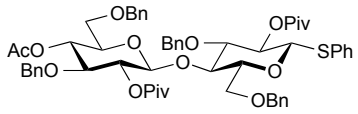
***p*-hydroxybenzhydrazide (pHBH) method.** Reducing sugar content was determined as described by Moretti et al.[189] The avicelase supernatant (40 μL) was diluted with water (60 μL) and combined with a freshly prepared 1:1 mixture (100 μL) of reagent A (2 M NaOH) and reagent B (2 wt% 4-hydroxybenzhydrazide in 0.5 M HCl) in a 96-well plate. The plate was incubated at 75 °C for 30 min, cooled to rt, and analyzed for light absorbance at 410 nm. Reducing sugar was quantified in reference to a cellobiose standard curve.

5.5 Efforts towards the synthesis of a fluorogenic cellohexaoside

Thiophenyl 4-O-acetyl-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranoside (3.2): *Method 1*: Donor 3.**7** (5.8 g, 9.74 mmol) containing DTBMP (3.4 g, 16.5 mmol) and acceptor **3.6** (3.5 g, 6.53 mmol) containing DTBMP (3.4 g, 16.5 mmol) were separately dissolved in CH_2Cl_2 and toluene and concentrated under reduced pressure and stored at 150 mTorr over night. The donor mixture was dissolved in anhydrous CH_2Cl_2 (300 mL) and added to an oven-dried rb flask containing powdered 4Å molecular sieves (22 g) by cannulation under N_2 pressure. The acceptor solution was dissolved in anhydrous CH_2Cl_2 (250 mL) and charged with pelleted 4ÅMS (15 g). The mixtures were stirred for 1 hour at rt under N_2. The donor slurry was then cooled to -70°C over an acetone/CO_2 bath and held at this temperature for 10 min. A solution of Tf_2O in CH_2Cl_2 (1 M, 4.86 mL, 4.86 mmol) was added drop-wise down the side of the reaction flask to the donor mixture and the flask was warmed to -60°C ($CHCl_3/CO_2$) and held for 15 min. Meanwhile, the acceptor solution was cooled to -70°C and then added drop-wise to the donor flask over 2 hrs by cannulation under N_2 pressure. The flask was then warmed to -10°C over 5 hours. Consumption of the acceptor was monitored by TLC (R_f = 0.5; 30% EtOAc/ hexanes), and the reaction was quenched with Et_3N (5mL, 36mmol) then filtered through celite and dried onto SiO_2 (20 g) under reduced pressure. The crude was purified via flash chromatography over SiO_2 (220 g, 55 mL/min, gradient elution 0→15% EtoAc/ hexanes). Residual DTBMP (2.5 g, 37% recovery) was recovered by an initial elution with pure hexanes, and the major

byproduct (**3.6a**, 657 mg, 17.4% yield) was isolated at ~13% EtOAc/ hexanes. Desired product (4.02 g, 61.3% yield) was isolated as a white foam with 15% EtOAc/ hexanes (isocratic) followed by residual acceptor (**3.6**, 364 mg, 10% recovery).

Method 2: The title compound can also be obtained by acetylation of compound **3.11** as described below for the formation of compound **3.6a**.



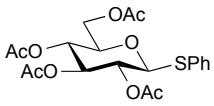
Chemical Formula: $C_{58}H_{68}O_{13}S$
Exact Mass: 1004.44
Molecular Weight: 1005.23

2

3.2: ^{1}H NMR (800 MHz, cdcl$_3$) δ 7.51 – 7.48 (m, 2H, arom Hs), 7.36 – 7.33 (m, 4H, arom Hs), 7.32 – 7.28 (m, 5H, arom Hs), 7.28 – 7.22 (m, 10H, arom Hs), 7.22 – 7.18 (m, 4H, arom Hs), 5.04 (t, *J* = 8.8 Hz, 1H, H^{II}-2), 5.04 (t, *J* = 9.7 Hz, 1H, H^{I}-2), 5.00 (t, *J* = 9.5 Hz, 1H, H^{II}-4), 4.97 (d, *J* = 11.0 Hz, 1H, PhC$\underline{H}_2$), 4.71 (d, *J* = 11.9 Hz, 1H, PhC$\underline{H}_2$), 4.62 (d, *J* = 10.1 Hz, 1H, H^{I}-1), 4.58 (d, *J* = 11.2 Hz, 2H, PhC$\underline{H}_2$), 4.52 (d, *J* = 11.1 Hz, 1H, PhC$\underline{H}_2$), 4.51 (d, *J* = 8.0 Hz, 1H, H^{II}-1), 4.49 (d, *J* = 11.9 Hz, 1H, PhC$\underline{H}_2$), 4.26 (d, *J* = 11.7 Hz, 1H, PhC$\underline{H}_2$), 4.13 (d, *J* = 11.7 Hz, 1H, PhC$\underline{H}_2$), 4.02 (t, *J* = 9.4 Hz, 1H, H^{I}-4), 3.77 – 3.75 (m, 2H, H^{II}-6, H^{II}-6'), 3.65 (t, *J* = 9.0 Hz, 1H, H^{I}-3), 3.55 (t, *J* = 9.4 Hz, 1H, H^{II}-3), 3.42 – 3.39 (m, 2H, H^{I}-5, H^{II}-5), 3.35 (dd, *J* = 10.8, 3.9 Hz, 1H, H^{I}-6), 3.19 (dd, *J* = 10.8, 5.7 Hz, 1H, H^{I}-6'), 1.81 (s, 3H, Ac-Hs), 1.19 (s, 9H, Piv-Hs), 1.16 (s, 9H, Piv-Hs). ^{13}C NMR (126 MHz, cdcl$_3$) δ 176.72 , 176.59 (Piv C=Os), 169.74 (Ac C=O), 138.85, 138.24, 138.12, 137.95, 133.40, 132.58, 128.97, 128.67,

128.49, 128.44, 128.21, 127.96, 127.91, 127.87, 127.80, 127.63, 127.50, 127.30 (arom Cs), 99.67 (C^{II}-1), 86.75 (C^{I}-1), 82.40 (C^{I}-3), 80.61 (C^{II}-3), 79.42 (C^{II}-5), 75.34 (C^{I}-4), 74.99 (Ph$\underline{C}H_2$), 73.79 (C^{I}-5, Ph$\underline{C}H_2$), 73.53 (Ph$\underline{C}H_2$), 72.79 (C^{I}-2), 71.20 (C^{II}-2), 71.02 (C^{II}-4), 70.04 (C^{I}-6), 68.23(C^{II}-6), 38.92, 38.85 (Piv 4°Cs), 27.42, 27.27 (Piv $\underline{C}H_3$), 20.92 (Ac $\underline{C}H_3$). ESI(+) HRMS calcd for $C_{58}H_{72}NO_{13}S$ ($[M+NH_4]^+$): 1022.4719; found: 1022.4732.

Thiophenyl 2,3,4,6-tetra-*O*-acetyl-*β*-D-glucopyranoside (3.3a): β-D-glucose pentaacetate (**3.3**, 30 g, 77 mmol) was dissolved in CH_2Cl_2 (300 mL) at 0°C and thiophenol (10.2 mL, 100 mmol) was added. After 30 min, BF_3•OEt_2 (29 mL, 231 mmol) was added drop-wise. The mix warmed to rt over night. After 16 hrs, the light pink solution was cooled to 0°C and the reaction quenched with saturated aqueous $NaHCO_3$ (200 mL) and the product was extracted into CH_2Cl_2 (500 mL). The extract was washed with $NaHCO_3$ (sat, 100 mL) and brine (150 mL) and dried with anhydrous Na_2SO_4. The solution was concentrated to a crude syrup which dried to a white solid (36 g) after azeotropic removal of moisture with toluene (3 x 20 mL). The crude was dissolved in hot toluene (30 mL) and re-crystallized as colorless crystals upon cooling to rt (29.7g , 88% yield).



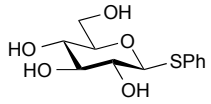
Chemical Formula: $C_{20}H_{24}O_9S$
Exact Mass: 440.11
Molecular Weight: 440.46

3a

3.3a: ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.54 – 7.45 (m, 2H, arom Hs), 7.38 – 7.25 (m, 3H, arom Hs), 5.23 (t, *J* = 9.4 Hz, 1H, H-3), 5.05 (t, *J* = 9.8 Hz, 1H, H-4), 4.98 (t, *J* = 9.7 Hz, 1H, H-2), 4.72 (d, *J* = 10.1 Hz, 1H, H-1), 4.23 (A of an ABX system, *J*= 12.3, 5.3 Hz, 1H, H-6), 4.19 (B of an ABX system, *J*= 12.3, 2.0 Hz, 1H, H-6'), 3.74 (X of an ABX system, apparent ddd, *J* = 10.0, 5.3, 2.0 Hz, 1H), 2.09 (s, 3H, Ac-Hs), 2.08 (s, 3H, Ac-Hs), 2.02 (s, 3H, Ac-Hs), 1.99 (s, 3H, Ac-Hs). ^{13}C NMR (100 MHz, cdcl$_3$) δ 170.52, 170.13, 169.38, 169.22, 133.10, 131.65, 128.95, 128.42, 85.67, 75.78, 73.96, 69.94, 68.22, 62.14, 20.76, 20.60. ESI(+) HRMS calcd for $C_{20}H_{28}NO_9S$ ($[M+NH_4]^+$): 458.1479; found: 1022.4732.

Thiophenyl 2,3,4,6-tetra-*O*-trimethylsilyl-*β*-D-glucopyranoside (3.4): Compound **3.3a** (15 g, 34 mmol) was dissolved in methanol (150 mL) and 25 wt% $NaOCH_3$ in methanol (0.74 mL, 3.4 mmol) was added. The mixed stirred at rt for 20 hrs, and DOWEX[H^+, 50W] (1.8 g) was added, resulting in a solution pH near 4 by wet pH paper. Activated charcoal was added and the slurry filtered through celite and concentrated to a white solid (10.9 g). The crude was dissolved in 1,4-dioxane (10 mL) and titrated while hot with toluene (15 mL) until cloudy. The mix cooled to rt, forming a white, crystalline solid (thiophenyl

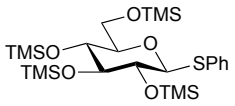
β-D-glucopyranoside (**3.3b**), 9.2 g, 99% yield). The solid was then dissolved in anhydrous pyridine (66 mL) at 0°C and TMSCl (21.7 mL, 170 mmol) was added drop-wise. The mix warmed to rt and stirred over night. The reaction was quenched by pouring over iced NH_4Cl (sat, 250 mL). The mixture was extracted with hexanes (3 x 200 mL), and the extracts washed with aqueous NH_4Cl (3 x 100 mL) and brine (200 mL) and then dried over anhydrous Na_2SO_4 before concentrating to an oil. The oil was concentrated in toluene on the rotary evaporator, and the product crystallized as colorless, glassy crystals during storage at 4°C (19 g, 98% yield).



Chemical Formula: $C_{12}H_{16}O_5S$
Exact Mass: 272.07
Molecular Weight: 272.32

3b

3.3b: ^{1}H NMR (500 MHz, d_2o) δ 7.49 (d, *J* = 7.9 Hz, 2H, arom Hs), 7.35 – 7.26 (m, 2H, arom Hs), 4.69 (d, *J* = 9.9 Hz, 1H, H-1), 3.81 (d, *J* = 12.4 Hz, 1H, H-6), 3.64 (dd, *J* = 12.5, 5.3 Hz, 1H, H-6'), 3.45 (t, *J* = 8.8 Hz, 1H, H-3), 3.40 – 3.31 (m, 2H, H-4, H-5), 3.28 (t, *J* = 9.4 Hz, 1H, H-2). ^{13}C NMR (100 MHz, d_2o) δ 132.02, 131.58, 129.32, 128.07, 87.26, 79.84, 77.21, 71.72, 69.33, 60.79. ESI(+) HRMS calcd for $C_{12}H_{16}NaO_5S$ ($[M+Na]^+$): 295.0611; found: 295.0610.



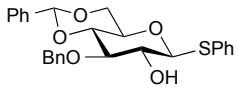
Chemical Formula: $C_{24}H_{48}O_5SSi_4$
Exact Mass: 560.23
Molecular Weight: 561.04

4

3.4: ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.54 (d, *J* = 8.2 Hz, 2H, arom Hs), 7.30 – 7.24 (m, 2H, arom Hs), 7.24 – 7.19 (m, 1H, arom H), 4.65 (d, *J* = 8.4 Hz, 1H, H-1), 3.83 (dd, *J* = 11.1, 1.3 Hz, 1H, H-6), 3.65 (dd, *J* = 11.1, 6.4 Hz, 1H, H-6'), 3.53 – 3.42 (m, 2H, H-3, H-4, H-5), 3.32 (t, *J* = 6.0 Hz, 1H, H-2), 0.26 (s, 9H, SiC$\underline{H}_3$), 0.19 (s, 9H, SiC$\underline{H}_3$), 0.18 (s, 9H, SiC$\underline{H}_3$), 0.12 (s, 9H, SiC$\underline{H}_3$). ^{13}C NMR (100 MHz, cdcl$_3$) δ 135.87 (s), 130.56, 128.83, 126.78, 89.39, 81.42, 80.03, 75.24, 71.78, 62.75, 1.84, 1.52, 1.07, -0.15. ESI(+) HRMS calcd for $C_{24}H_{52}NO_5SSi_4$ ($[M+NH_4]^+$): 578.2638; found: 578.2636.

Thiophenyl 3-*O*-benzyl-4,6-*O*-benzylidene-*β*-D-glucopyranoside (3.5): Compound **3.4** (1.71 g, 3.04 mmol) was dissolved in CH_2Cl_2 (6.1 mL), combined with benzaldehyde (0.68 mL, 6.69 mmol), and cooled to 0°C. A freshly-prepared solution of newly-dried Cu(OTf)$_2$ in acetonitrile (74 mM, 1.42 mL, 0.106 mmol) was added drop-wise over 5 min and the mix warmed to rt over 30 min and then to 30°C. After 1 hr, TLC (9:1 Hexanes/EtOAc) indicated no remaining starting material, and the mixture was cooled to 0°C. Triethylsilane (0.53 mL, 3.34 mmol) was added drop-wise, and the mix stirred for 45 min at 0°C until the newly formed intermediate was consumed as monitored by TLC (Rf =

0.6, 4:1 hexanes/ EtOAc). The solution often became pink during this process. Then a TBAF solution in THF (1 M, 6.1 mL, 6.1 mmol) and solid $NaHCO_3$ (510 mg, 6.1 mmol) was added to the reaction, and mix (now orange) warmed to rt. After 30 min, the mixture was diluted with CH_2Cl_2 (300 mL) and washed with aqueous 10% $NaHCO_3$ (50 mL), water (2 x 50 mL), and brine (60 mL). The solution was dried over Na_2SO_4 and concentrated to a tan powder (2 g). The crude was dissolved in hot ethanol (3 mL) and cooled to rt, to yield product as white, feathery crystals (983 mg, 72% yield). More product was isolated by flash chromatography by elution over silica gel to 9:1 hexanes/ EtOAc (186 mg, 13.6%).

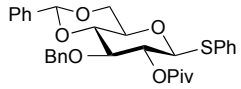


Chemical Formula: $C_{26}H_{26}O_5S$
Exact Mass: 450.15
Molecular Weight: 450.55

5

3.5: ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.52 – 7.41 (m, 4H, arom Hs), 7.38 – 7.21 (m, 11H, arom Hs), 5.52 (s, 1H, PhC<u>H</u>(OR)$_2$), 4.91 (d, *J* = 11.6 Hz, 1H, PhC<u>H</u>$_2$), 4.75 (d, *J* = 11.6 Hz, 1H, PhC<u>H</u>$_2$), 4.59 (d, *J* = 9.7 Hz, 1H, H-1), 4.34 (dd, *J* = 10.5, 5.0 Hz, 1H, H-6), 3.75 (t, *J* = 10.3 Hz, 1H, H-4), 3.67 – 3.58 (m, 2H, H-3, H-6'), 3.51 – 3.42 (m, 2H, H-2, H-5), 2.59 (d, *J* = 2.2 Hz, 1H, OH). ^{13}C NMR (126 MHz, cdcl$_3$) δ 138.27, 137.28, 133.24, 131.45, 129.13, 128.56, 128.43, 128.36 (2C), 128.20, 127.97, 126.10, 101.32, 88.53, 81.71, 81.18, 74.89, 72.33, 70.80, 68.71. ESI(+) HRMS calcd for $C_{26}H_{26}NaO_5S$ ($[M+Na]^+$): 473.1393; found: 473.1394.

Thiophenyl 3-*O*-benzyl-4,6-*O*-benzylidene-2-*O*-pivaloyl-*β*-D-glucopyranoside (3.5a): Alcohol **3.5** (6.74 g, 12 mmol), Et_3N (8.3 mL, 60 mmol), and DMAP (1.46 g, 12 mmol) were combined in CH_2Cl_2 (40 mL) and cooled to 0°C. PivCl (5.9 mL, 48 mmol) was added and the mix warmed to rt. After 5.5 hrs, reaction mixture was poured over iced 10% $NaHCO_3$ (200 mL) and extracted with CH_2Cl_2 (3 x 100 mL). The extracts were washed with 10% K_2CO_3 (2 x 100 mL), 10% $NaHCO_3$ (2 x 100 mL), water (2 x 100 mL), and brine (150 mL). The solution was dried over anhydrous Na_2SO_4 and concentrated to a powder. The powder was dissolved in hot ethanol upon titration with chloroform and recrystallized upon cooling to rt to yield pure product (6.19 g, 97%) as white crystals.

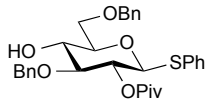


Chemical Formula: $C_{31}H_{34}O_6S$
Exact Mass: 534.21
Molecular Weight: 534.67

5a

3.5a: ^1H NMR (500 MHz, cdcl$_3$) δ 7.48 – 7.41 (m, 4H, arom Hs), 7.39 – 7.17 (m, 11H, arom Hs), 5.54 (s, 1H, PhC<u>H</u>(OR)$_2$), 5.07 (dd, *J* = 9.5, 8.8 Hz, 1H, H-2), 4.85 (d, *J* = 11.5 Hz, 1H, PhC<u>H</u>$_2$), 4.73 (d, *J* = 10.1 Hz, 1H, H-1), 4.63 (d, *J* = 11.5 Hz, 1H, PhC<u>H</u>$_2$), 4.37 (dd, *J* = 10.5, 5.0 Hz, 1H, H-6), 3.82 – 3.70 (m, 3H, H-3, H-4, H-6′), 3.50 (apparent td, *J* = 9.5, 5.0 Hz, 1H, H-5), 1.21 (s, 9H, Piv-Hs). ^{13}C NMR (126 MHz, cdcl$_3$) δ 176.61, 138.11, 137.25, 132.78, 132.67, 129.12, 129.06, 128.36, 128.25, 127.67, 127.58, 126.10, 101.32, 87.36, 81.31, 80.50, 74.63, 71.07, 70.63, 68.67, 38.84, 27.26. ESI(+) HRMS calcd for $C_{31}H_{38}NO_6S$ ($[M+NH_4]^+$): 552.2414; found: 552.2414.

Thiophenyl 3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranoside (3.6): Compound **3.5a** (4.75 g, 8.9 mmol) was combined with sodium cyanoborohydride (5.59 g, 89 mmol) and 4Å molecular sieves (1.5 g) in anhydrous THF (120 mL). The slurry was cooled to 0°C and dry HCl in dioxane (4 M, 22.3 mL, 89 mmol) was added drop-wise over 2 hrs, resulting in pH ~ 2-3. The mixture was warmed to rt over 30 min and stirred at rt for 30 min until no starting material remained by TLC (4:1 hexanes/ EtOAc). The mixture was filtered through celite into ice-cold, aqueous 10% $NaHCO_3$ (75 mL) and then extracted into EtOAc (3 x 100 mL). The extracts were washed with water (3 x 100 mL) and brine (120 mL) then dried over Na_2SO_4 and concentrated under reduced pressure onto silica gel (15 g). The product was isolated by flash chromatography by elution to 15% EtoAc/ hexanes over SiO_2 (24 g, 35 ml/min) to yield desired product as a colorless foam (4.33 g, 91% yield).



Chemical Formula: $C_{31}H_{36}O_6S$
Exact Mass: 536.22
Molecular Weight: 536.68

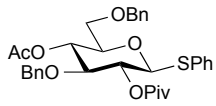
6

3.6: ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.38 – 7.32 (m, 4H, arom Hs), 7.23 – 7.07 (m, 11H, arom Hs), 4.92 (t, *J* = 9.6 Hz, 1H, H-2), 4.58, 4.54 (ABq, *J* = 11.3 Hz, 2H, PhC$\underline{H}_2$), 4.52 (d, *J* = 10.1 Hz, 1H, H-1), 4.43, 4.41 (ABq, *J* = 11.9 Hz, 2H, PhC$\underline{H}_2$), 3.65 (A of an ABX system, *J* = 10.5, 3.7 Hz, 1H, H-6), 3.61 (B of an ABX system, *J* = 10.5, 5.5 Hz, 1H, H-6'), 3.56 (t, *J* = 9.3 Hz,

1H, H-3), 3.43 (t, *J* = 9.0 Hz, 1H, H-4), 3.39 (X of ABX, apparent m, 1H, H-5), 2.62 (s, 1H, OH), 1.10 (s, 9H, Piv-Hs). ^{13}C NMR (126 MHz, cdcl$_3$) δ 176.88, 138.15, 137.95, 133.53, 132.26, 128.99, 128.65, 128.55, 127.98, 127.89, 127.83, 127.66, 86.92, 84.31, 78.50, 74.80, 73.80, 71.51, 71.25, 70.37, 38.91, 27.30.

Phenyl (4-O-acetyl-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl)sulfoxide (3.7a & 3.7b): Compound **3.6** (5.22 g, 9.74 mmol) was combined with Et$_3$N (2 mL, 14.6 mmol) and DMAP (118 mg, 0.97 mmol) in CH_2Cl_2 (75 mL) and the mixture was stirred in open atmosphere at rt. Acetic anhydride (1.1 mL, 11.7 mmol) was added and the mixture stirred for 1:45 hrs. The mix was diluted with CH_2Cl_2 (400 mL), and washed with 10% $NaHCO_3$ (100 mL), water (3 x 100 mL), and brine (100 mL). The solution was dried over Na_2SO_4, and the solvents were removed under reduced pressure to yield pure thiophenyl 4-*O*-acetyl-3,6-*O*-benzyl-2-*O*-pivaloyl-β-D-glucopyranoside (**3.6a**, 5.6 g, 100%) as a white solid. The solid was dissolved in anhydrous CH_2Cl_2 (160 mL) and cooled to -10°C with iced brine. After 10 min, solid mCPBA (~75%, 2.4 g, 10.7 mmol) was added to the mixture and the mixture stirred in open atmosphere for 45 min while warming to -5°C. When starting material (**3.6a**) had been consumed on TLC (Rf = 0.6; 30% EtOAc/ hexanes) the reaction was poured over iced 10% $NaHCO_3$ (100 mL) and extracted into CH_2Cl_2 (300 mL). The extracts were washed with 10% $NaHCO_3$ (100 mL), 1M $Na_2S_2O_3$ (50 mL), 5% $NaHCO_3$ (100 mL), and brine (150 mL). The solution was dried

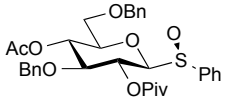
over Na_2SO_4 and co-evaporated with toluene under reduced pressure to yield the title compound (5.67 g, waxy white solid, 98%) as a mixture of sulfoxide diasteromers.



Chemical Formula: $C_{33}H_{38}O_7S$
Exact Mass: 578.23
Molecular Weight: 578.72

6a

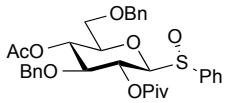
3.6a: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.53 – 7.48 (m, 1H, arom Hs), 7.37 – 7.17 (m, 7H, arom Hs), 5.13 (dd, *J* = 10.0, 9.2 Hz, 1H, H-2), 5.06 (t, *J* = 9.5 Hz, 1H, H-4), 4.70 (d, *J* = 10.1 Hz, 1H, H-1), 4.65 (d, *J* = 11.2 Hz, 1H, PhC$\underline{H}_2$), 4.56 – 4.49 (m, 3H, PhC$\underline{H}_2$), 3.77 (t, *J* = 9.2 Hz, 1H, H-3), 3.70 – 3.55 (m, 3H, H-5,6,6'), 1.85 (s, 3H, Ac-Hs), 1.25 (s, 9H, Piv-Hs). ^{13}C NMR (75 MHz, cdcl$_3$) δ 176.64 (Piv C=O), 169.71 (Ac C=O), 138.04, 137.91 (4° arom Cs), 133.44 (C^{SPh} 4° arom), 132.24, 129.04, 128.49, 127.94, 127.90, 127.81, 127.57 (arom Cs), 86.87 (C-1), 81.98 (C-3), 77.97 (C-5), 73.97, 73.75 (Ph$\underline{C}H_2$), 71.07, 70.48, 69.94 (C-2,4,6), 38.90 (4° Piv C), 27.30 (Piv $\underline{C}H_3$), 20.89 (Ac $\underline{C}H_3$). ESI(+) HRMS calcd for $C_{33}H_{42}NO_7S$ ($[M+NH_4]^+$): 596.2676; found: 596.2675.



Chemical Formula: $C_{33}H_{38}O_8S$
Exact Mass: 594.23
Molecular Weight: 594.72

7a

3.7a: (less polar sulfoxide) ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.63 – 7.59 (m, 2H, arom Hs), 7.43 – 7.39 (m, 3H, arom Hs), 7.33 – 7.21 (m, 7H, arom Hs), 7.17 (t, *J* = 7.2 Hz, 3H, arom Hs), 5.45 (t, *J* = 9.3 Hz, 1H, H-2), 4.99 (t, *J* = 9.2 Hz, 1H, H-4), 4.61, 4.53 (ABq, *J* = 11.1 Hz, 2H, PhC$\underline{H}_2$), 4.30, 4.26 (ABq, *J* = 11.7 Hz, 2H, PhC$\underline{H}_2$), 4.20 (d, *J* = 9.7 Hz, 1H, H-1), 3.84 (t, *J* = 9.0 Hz, 1H, H-3), 3.59 – 3.42 (m, 3H, H-5, H-6, H-6′), 1.81 (s, 3H, Ac-Hs), 1.24 (s, 9H, Piv-Hs). ^{13}C NMR (126 MHz, cdcl$_3$) δ 176.24, 169.54, 139.44, 137.91, 131.59, 129.04, 128.52, 128.45, 127.89, 127.83, 127.66, 125.65, 91.08, 81.39, 79.07, 73.70, 73.40, 69.75, 68.18, 29.84, 27.27, 20.82. ESI(+) HRMS calcd for $C_{33}H_{38}NaO_8S$ ([M+Na]$^+$): 617.2180; found: 617.2181.

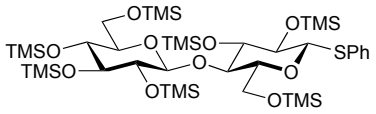


Chemical Formula: $C_{33}H_{38}O_8S$
Exact Mass: 594.23
Molecular Weight: 594.72

7b

3.7b: (more polar sulfoxide) ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.76 – 7.67 (m, 2H, arom Hs), 7.46 – 7.33 (m, 4H, arom Hs), 7.29 – 7.09 (m, 9H, arom Hs), 5.09 (t, *J* = 8.8 Hz, 1H, H-2), 4.84 (t, *J* = 9.1 Hz, 1H, H-4), 4.61 (d, *J* = 11.3 Hz, 1H, PhC$\underline{H}_2$), 4.49 (d, *J* = 11.3 Hz, 1H, PhC$\underline{H}_2$), 4.38 (d, *J* = 9.4 Hz, 1H, H-1), 4.27 (ABq, *J* = 11.8 Hz, 2H, PhC$\underline{H}_2$), 3.72 (t, *J* = 8.4 Hz, 1H, H-3), 3.58 (Apparent dt, *J* = 9.4, 4.3 Hz, 1H, H-5), 3.43 – 3.34 (m, 2H, H-6, H-6′), 1.75 (s, 3H, Ac-Hs), 1.18 (s, 9H, Piv-Hs). ^{13}C NMR (126 MHz, cdcl$_3$) δ 176.93, 169.44, 138.48, 137.90, 137.58, 131.93, 128.81, 128.55, 128.46, 127.96, 127.86, 127.82, 127.61, 126.62, 93.43, 81.20, 78.10, 74.24, 73.82, 69.85, 69.23, 69.07, 39.04, 27.23, 20.83. ESI(+) HRMS calcd for $C_{33}H_{38}NaO_8S$ ([M+Na]$^+$): 617.2180; found: 617.2186.

Thiophenyl 2,3,4,6-O-tetra(trimethylsilyl)-β-D-glucopyranosyl-(1→4)-2,3,6-*O*-tri(trimethylsilyl)-β-D-glucopyranoside (3.8): Thiophenyl heptaacetyl-β-D-cellobioside is commercially available (Abacipharm Corporation, Columbia, MD, USA) or can be prepared from the biphasic substitution reaction with thiophenol and acetobromocellobiose (*vide infra*).[292-294] Thiophenyl heptaacetyl-β-D-cellobioside (7.56 g, 10.4 mmol) was stirred in methanol (75 mL) under Ar at rt. A solution of sodium methoxide in methanol (1.9 M, 2.73 mL, 5.19 mmol) was added drop-wise and the mixture stirred at rt for 16 hours. The reaction was quenched with acetic acid (1 mL) and co-rotovapped with toluene to yield a white foam. The foam was combined with imidazole (707 mg, 10.4 mmol) and dissolved in DMF (75 mL) at rt under Ar. Et_3N (29 mL, 207.6 mmol) was added and the mixture was cooled to 0°C. TMSCl (21 mL, 166 mmol) was added with an addition funnel over 30 min. The reaction was complete after 2.5 hrs at rt. The mixture was poured over iced sodium bicarbonate (~5%, 300 mL) and extracted with Et_2O (3 X 200 mL). The extracts were washed with 10% $NaHCO_3$ (200 mL), water (3 X 200 mL) and brine (200 mL) then dried over $MgSO_4$. The ether was removed under vacuum, and the crude reside was dissolved in minimal EtOAc and recrystallized from hot EtOH (30 mL). The title compound was obtained as colorless crystals (8.98 g, 92% yield).



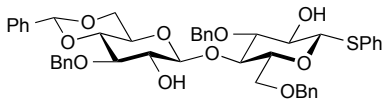
Chemical Formula: $C_{39}H_{82}O_{10}SSi_7$
Exact Mass: 938.40
Molecular Weight: 939.73

8

3.8: ^{1}H NMR (400 MHz, cdcl$_3$) δ 7.55 – 7.49 (m, 2H, arom Hs), 7.30 – 7.19 (m, 3H, arom Hs), 4.55 (d, *J* = 9.0 Hz, 1H, H^{I}-1), 4.40 (d, *J* = 7.3 Hz, 1H, H^{II}-1), 3.89 – 3.75 (m, 4H, H^{I}-4,6,6', H^{II}-6), 3.59 (dd, *J* = 10.9, 5.8 Hz, 1H, H^{II}-6'), 3.54 – 3.44 (m, 2H, H^{I}-2,3), 3.36 – 3.19 (m, 4H, H^{I}-5, H^{II}-2,3,4), 3.17 – 3.09 (m, 1H, H^{II}-5), 0.22 (s, 9H, Si-CHs), 0.18 – 0.11 (m, 54H, Si-CHs). ^{13}C NMR (101 MHz, cdcl$_3$) δ 135.27, 131.39, 128.83, 127.05 [arom Cs], 101.41 (C^{II}-1), 89.39 (C^{I}-1), 80.21 (C^{I}-5), 78.68 (C^{I}-3), 78.50 (C^{II}-3), 77.69 (C^{II}-5), 75.77 (C^{I}-2), 75.18 (C^{II}-2), 73.87 (C^{I}-4), 72.04 (C^{II}-4), 62.49 (C^{II}-6), 61.27 (C^{I}-6), 1.71, 1.67, 1.51, 1.49, 1.18, 0.01 (2C) [Si-Cs].

Thiophenyl 3-*O*-benzyl-4,6-*O*-benzylidene-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-*β*-D-glucopyranoside (3.9): Compound **3.8** (5.91 g, 6.29 mmol) was dissolved in anhydrous dichloromethane (25 mL) under Ar. Benzaldehyde (2.8 mL, 27.68 mmol) was added and the mixture cooled to 0°C. A freshly prepared solution of copper(II) trifluoromethanesulfonate (137 mg, 0.377 mmol) in anhydrous acetonitrile (6.3 mL) was added drop-wise over 5 min. The mix stirred for 1 hr at 0°C and 30 min at rt, at which point, the starting material was consumed as monitored by TLC (10% EtOAc/ hexanes). The mixture was cooled again to 0°C, and triethylsilane (3.1 mL, 19.5 mmol) was added

drop-wise using an addition funnel over 5—10 min. The mixture stirred at 0°C for 1 hour, and solid sodium bicarbonate (2.32 g, 27.68 mmol) was added. A solution of tetrabutylammonium fluoride in THF (1 M, 27.7 mL, 27.68 mmol) was added and the mixture stirred while warming to rt over 2.5 hours. The mixture was diluted with dichloromethane (600 mL) and washed with water (3 X 200 mL) and brine (200 mL). The organic layer was dried over Na_2SO_4 and concentrated under vacuum onto SiO_2 (8 g). The product was purified via flash chromatography (125 g SiO_2, 0—20% EtoAc/ hexanes) and isolated as a colorless foam. The foam was crystallized from EtOAc (5 mL) and hot EtOH (35 mL), yielding pure **3.9** (2.08 g, 41.8% yield) as colorless crystals.



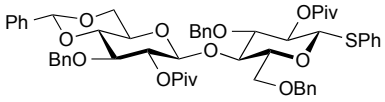
Chemical Formula: $C_{46}H_{48}O_{10}S$
Exact Mass: 792.30
Molecular Weight: 792.94

9

3.9: ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.55 – 7.51 (m, 2H, arom Hs), 7.46 – 7.42 (m, 2H, arom Hs), 7.37 – 7.20 (m, 21H, arom Hs), 5.44 (s, 1H, PhC<u>H</u>), 4.93 – 4.81 (m, 2H, PhC<u>H</u>$_2$), 4.76 (d, *J* = 11.7 Hz, 1H, PhC<u>H</u>$_2$), 4.64 (dd, *J* = 16.8, 9.6 Hz, 2H, PhC<u>H</u>$_2$, H^{II}-1), 4.55 (d, *J* = 11.9 Hz, 1H, PhC<u>H</u>$_2$), 4.45 (d, *J* = 9.6 Hz, 1H, H^{I}-1), 4.02 – 3.90 (m, 3H, H^{I}-6,6', H^{II}-4), 3.80 (d, *J* = 10.4 Hz, 1H, H^{II}-6), 3.60 – 3.39 (m, 7H, H^{I}-2-5, H^{II}-2,3,4,6'), 3.33 (s, 1H, OH), 3.13 (apparent td, *J* = 9.5, 5.0 Hz, 1H, H^{II}-5), 2.63 (s, 1H, OH). ^{13}C NMR (126 MHz, cdcl$_3$) δ 138.79, 138.50, 137.81, 137.33 (4˚ arom Cs), 133.26 (arom C), 131.45 (arom S-C), 128.98, 128.46, 128.41, 128.24, 127.98, 127.95, 127.88, 127.75, 127.65, 127.26, 126.07 (arom Cs),

103.60 (C^{II}-1), 101.16 (PhCH), 87.79 (C^{I}-1), 84.74 (C^{I}-5), 81.25 (C^{II}-3), 80.38 (C^{I}-3), 79.07 (C^{II}-4), 76.85 (C^{II}-2), 75.18 (C^{I}-4), 74.96 ($PhCH_2$), 74.54 ($PhCH_2$), 73.62 ($PhCH_2$), 72.23 (C^{I}-2), 68.62 (C^{I}-6), 68.51 (C^{II}-6), 66.36 (C^{II}-5).

Thiophenyl 3,-*O*-benzyl-4,6-*O*-benzylidene-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranoside (3.10): Compound **3.9** (1.21 g, 1.53 mmol) was dissolved in dry acetonitrile (6.1 mL) in a flame-dried flask under Ar. Pivalic anhydride (???) was added. The solution was cooled to 0°C and a solution of scandium(III) trifluoromethanesulfonate in acetonitrile (0.1 M, 0.77 mL, 0.077 mmol) was added drop-wise. The reaction stirred for 3 hrs, warming to 4°C until completion. The mixture was poured into 10% $NaHCO_3$ (75 mL) and extracted with EtOAc (3 X 100 mL). The extracts were washed with water (2 X 75 mL) and brine (50 mL) then dried over Na_2SO_4. The solvents were removed under vacuum, and the product was recrystallized from hot EtOH (25 mL), forming white, fibrous crystals (1.16 g, 79% yield).



Chemical Formula: $C_{56}H_{64}O_{12}S$
Exact Mass: 960.41
Molecular Weight: 961.18

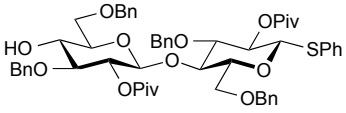
10

3.10: ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.51 – 7.48 (m, 2H, arom Hs), 7.45 – 7.42 (m, 2H, arom Hs), 7.39 – 7.21 (m, 21H, arom Hs), 5.40 (s, 1H, PhCH), 5.06 – 4.97 (m, 2H, H^{I}-2, H^{II}-2), 4.91 (d, *J* = 11.0 Hz, 1H, $PhCH_2$), 4.87 (d, *J* = 11.6 Hz, 1H, $PhCH_2$), 4.72 (d, *J* = 11.9 Hz, 1H,

PhC$\underline{H}_2$), 4.64 – 4.52 (m, 4H, H^{I}-1, H^{II}-1, PhC$\underline{H}_2$), 4.47 (d, *J* = 11.9 Hz, 1H, PhC$\underline{H}_2$), 4.14 – 4.05 (m, 2H, H^{I}-4, H^{II}-6), 3.81 – 3.73 (m, 2H, H^{I}-6,6'), 3.67 – 3.54 (m, 3H, H^{I}-3, H^{II}-3,4), 3.39 (bd, *J* = 9.7 Hz, 1H, H^{I}-5), 3.27 (t, *J* = 10.2 Hz, 1H, H^{II}-6'), 3.23 – 3.17 (m, 1H, H^{II}-5), 1.20 (s, 9H, Piv-Hs), 1.15 (s, 9H, Piv-Hs). ^{13}C NMR (126 MHz, cdcl$_3$) δ 176.64, 176.56 (C=0s), 138.81, 138.18, 137.98, 137.30, 133.00 (4° arom Cs), 132.75, 129.03, 128.88, 128.61, 128.28, 128.15, 128.08, 128.01, 127.93, 127.56, 127.40, 127.30, 127.22, 126.06 (arom Cs), 101.17 (Ph$\underline{C}$H), 100.00 (C^{II}-1), 86.50 (C^{I}-1), 82.18 (C^{I}-4), 81.74 (C^{II}-4), 79.27 (C^{I}-5), 79.09 (C^{II}-3), 75.11 (C^{I}-4), 74.81, 74.21, 73.72 (Ph$\underline{C}H_2$), 73.16 (C^{II}-2), 70.90 (C^{I}-2), 68.50 (C^{II}-6), 68.06 (C^{I}-6), 66.02 (C^{II}-5), 38.81, 38.76 (Piv 4°Cs), 27.31, 27.20 (Piv$\underline{C}H_3$).

Thiophenyl 3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranoside (3.11): *Method 1*: Disaccharide **3.2** (437 mg, 0.44 mmol) was dissolved in a mixture of ethanol and CH_2Cl_2 (7:1, 2.2 mL) and an equal volume of guanidine (200 mM, 2.2 mL, 0.44 mmol) in the same ethanol/CH_2Cl_2 mixture was added. Methanolic $NaOCH_3$ (1.9 M, 47 μL, 0.09 mmol) was added and the mixture was stirred in open atmosphere for 22 hrs. The mixture was diluted with CH_2Cl_2 (250 mL) and washed with 10% $NaHCO_3$ (50 mL), water (3 x 50 mL), and brine (60 mL) and then dried over anhydrous $MgSO_4$. The solution was concentrated under reduced pressure to afford pure product as a white foam (381 mg, 91% yield).

Method 2: Sodium cyanoborohydride (111 mg, 1.77 mmol) was combined with 4Å MS (200 mg) in a flame-dried flask under Ar. A solution of compound **3.10** (170 mg, 0.177 mmol) in anhydrous THF (5 mL) was added to the reaction flask and the mixture was stirred at rt for 45 min. The reaction was cooled to 0°C and a solution of HCl in dioxane (4 M, 0.44 mL, 1.77 mmol) was added drop-wise over 45 min, resulting in a solution pH of ~2.5. The mixture slowly warmed to rt over 3 hrs, at which point no SM remained by TLC (30% EtOAc/ hexanes). Solid $NaHCO_3$ (300 mg) was added to quench the acid, and the mixture was filtered through celite. The filtrate was diluted with EtOAc (150 mL) and washed with 10% $NaHCO_3$ (50 mL), water (3 X 50 mL), and brine (50 mL). The organic layer was dried over Na_2SO_4 and concentrated onto SiO_2 (500 mg) under vacuum. The product was purified by flash chromatography (12 g SiO_2, 0 –20% EtOAc/ hexanes) yielding **3.11** (147 mg, 86% yield) as a white foam.



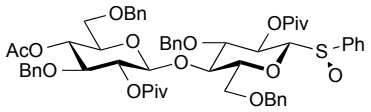
Chemical Formula: $C_{56}H_{66}O_{12}S$
Exact Mass: 962.43
Molecular Weight: 963.19

11

3.11: ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.44 – 7.41 (m, 2H, arom Hs), 7.32 – 7.09 (m, 23H, arom Hs), 4.96 (t, *J* = 9.6 Hz, 1H, H^{I}-2), 4.88 (t, *J* = 8.9 Hz, 1H, H^{II}-2), 4.87 (d, *J* = 11.2 Hz, 1H, PhC$\underline{H}_2$), 4.75 (d, *J* = 11.4 Hz, 1H, PhC$\underline{H}_2$), 4.66 (d, *J* = 11.9 Hz, 1H, PhC$\underline{H}_2$), 4.56 (d, *J* = 11.4 Hz, 1H, PhC$\underline{H}_2$), 4.55 (d, *J* = 10.1 Hz, 1H, H^{I}-1), 4.48 (d, *J* = 11.2 Hz, 1H, PhC$\underline{H}_2$), 4.43 –

4.37 (m, 2H, H^{II}-1, PhC<u>H</u>$_2$), 4.26, 4.22 (ABq, *J* = 11.9 Hz, 2H, PhC<u>H</u>$_2$), 3.95 (t, *J* = 9.4 Hz, 1H, H^{I}-4), 3.69 (s, 2H, H^{I}-6,6'), 3.60 (bt, *J* = 8.9 Hz, 1H, H^{II}-4), 3.54 (t, *J* = 9.0 Hz, 1H, H^{I}-3), 3.39 (X of an ABX system, apparent td, *J* = 7.1, 3.3 Hz, 1H, H^{II}-6), 3.34 – 3.27 (m, 2H, H^{I}-5, H^{II}-3), 3.22 (AB of an ABX system, apparent td, *J* = 10.6, 6.4 Hz, 2H, H^{II}-5,6'), 2.95 (bs, 1H, RO<u>H</u>), 1.11 (s, 9H, Piv C<u>H</u>$_3$), 1.08 (s, 9H, Piv C<u>H</u>$_3$). ^{13}C NMR (75 MHz, cdcl$_3$) δ 176.83, 176.71 (Piv C=Os), 138.96, 138.47, 138.04, 137.60 (4° arom Cs), 133.30 (C^{SPh} 4°C), 132.63, 128.95, 128.67, 128.61, 128.49, 128.19, 128.10, 128.02, 127.95, 127.92, 127.79, 127.69, 127.41, 127.26, 127.15 (arom Cs), 99.61 (C^{II}-1), 86.66 (C^{I}-1), 82.59 (C^{II}-3), 82.29 (C^{I}-3), 79.37 (C^{I}-5), 75.10 (C^{I}-4), 74.78, 74.44 (Ph<u>C</u>H$_2$), 74.18 (C^{II}-4), 73.82, 73.78 (Ph<u>C</u>H$_2$), 73.08 (C^{II}-5), 72.88 (C^{II}-2), 71.24 (C^{II}-6), 70.91 (C^{I}-2), 68.18 (C^{I}-6), 38.90, 38.82 (4° Piv Cs), 27.40, 27.24 (Piv <u>C</u>H$_3$). ESI(+) HRMS calcd for $C_{56}H_{70}NO_{12}S$ ([M+NH$_4$]$^+$): 980.4613; found: 980.4617.

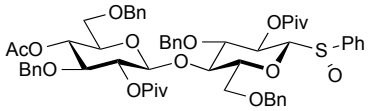
Phenyl (4-*O*-acetyl-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl)sulfoxide (3.12): The title compound (989 mg, 95% yield) was prepared as a white powder from disaccharide **3.2** (1.03 g, 1.02 mmol) and mCPBA (77%, 251 mg, 1.12 mmol) in a similar manner as compound **3.7**.



Chemical Formula: $C_{58}H_{68}O_{14}S$
Exact Mass: 1020.43
Molecular Weight: 1021.23

12a

3.12a: (less polar sulfoxide) ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.84 (d, *J* = 7.6 Hz, 2H, arom Hs), 7.44 (t, *J* = 7.5 Hz, 2H, arom Hs), 7.38 (dd, *J* = 13.3, 6.1 Hz, 2H, arom Hs), 7.35 – 7.23 (m, 14H, arom Hs), 7.23 – 7.19 (m, 3H, arom Hs), 7.15 (d, *J* = 6.9 Hz, 2H, arom Hs), 5.07 – 4.99 (m, 4H, H^{I,II}-2, H^{II}-4, PhC$\underline{H}_2$), 4.64 (d, *J* = 11.0 Hz, 1H, PhC$\underline{H}_2$), 4.59 (d, *J* = 11.2 Hz, 1H, PhC$\underline{H}_2$), 4.56 – 4.47 (m, 3H, H^{I}-1, H^{II}-1, PhC$\underline{H}_2$), 4.26 (d, *J* = 11.7 Hz, 1H, PhC$\underline{H}_2$), 4.19 – 4.12 (m, 3H, PhC$\underline{H}_2$), 3.92 (t, *J* = 9.0 Hz, 1H, H^{I}-4), 3.82 – 3.75 (m, 2H, H^{I}-3, H^{I}-6), 3.70 (d, *J* = 12.0 Hz, 1H, H^{I}-6′), 3.58 (t, *J* = 9.4 Hz, 1H, H^{II}-3), 3.46 (bd, *J* = 8.8 Hz, 1H, H^{I}-5), 3.39 – 3.33 (m, 2H, H^{II}-5, H^{I}-6), 3.21 (dd, *J* = 11.2, 6.1 Hz, 1H, H^{II}-6′), 1.80 (s, 3H, Ac Hs), 1.25 (s, 9H, Piv Hs), 1.16 (s, 9H, Piv Hs). ^{13}C NMR (126 MHz, cdcl$_3$) δ 177.11, 176.54 (Piv C=Os), 169.68 (Ac C=O), 138.45, 138.15, 137.90 , 131.76, 128.75, 128.54, 128.46, 128.39, 128.28, 127.85, 127.78, 127.70, 127.61, 127.58, 127.50, 127.46, 126.63 (arom Cs), 99.65 (C^{II}-1), 94.05 (C^{I}-1), 81.66 (C^{I}-3), 80.54 (C^{II}-3), 79.95 (C^{I}-5), 75.18 (Ph$\underline{C}H_2$), 74.54 (C^{I}-4), 73.84 (Ph$\underline{C}H_2$), 73.68 (C^{II}-5), 73.61 (Ph$\underline{C}H_2$), 72.67 (C^{II}-2), 71.19 (C^{I}-2), 69.96 (C^{II}-6), 68.67 (C^{II}-4), 67.62 (C^{I}-6), 39.00, 38.89 (Piv 4˚Cs), 27.34, 27.18 (Piv $\underline{C}H_3$), 20.86 (Ac $\underline{C}H_3$). ESI(+) HRMS calcd for $C_{58}H_{68}NaO_{14}S$ ([M+Na]$^{+}$): 1043.4222; found: 1043.4221.



Chemical Formula: $C_{58}H_{68}O_{14}S$
Exact Mass: 1020.43
Molecular Weight: 1021.23

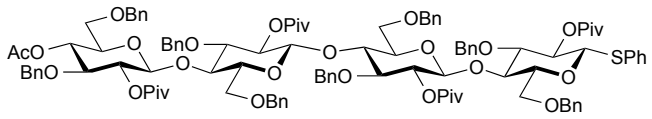
12b

3.12b: (more polar sulfoxide) ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.63 (dd, *J* = 7.2, 1.9 Hz, 2H, arom Hs), 7.48 – 7.14 (m, 23H, arom Hs), 5.42 (t, *J* = 9.1 Hz, 1H, H^{I}-2), 5.03 (dd, apparent

t, *J* = 9.3, 8.3 Hz, 1H, H^{II}-2), 4.98 (t, *J* = 9.6 Hz, 1H, H^{II}-4), 4.95 (d, *J* = 11.2 Hz, 1H, PhC$\underline{H}_2$), 4.63 – 4.54 (m, 3H, H^{I}-1, PhC$\underline{H}_2$), 4.50 (bd, *J* = 9.3 Hz, 2H, H^{II}-1, PhC$\underline{H}_2$), 4.38 (d, *J* = 12.0 Hz, 1H, PhC$\underline{H}_2$), 4.24 (d, *J* = 11.7 Hz, 1H, PhC$\underline{H}_2$), 4.15 (d, *J* = 11.8 Hz, 1H, PhC$\underline{H}_2$), 4.13 (d, *J* = 9.7 Hz, 1H, H^{I}-1), 4.00 (t, *J* = 8.9 Hz, 1H, H^{I}-4), 3.74 (t, *J* = 8.5 Hz, 1H, H^{I}-3), 3.67 (dd, *J* = 11.4, 4.5 Hz, 1H, H^{I}-6), 3.57 (m, 2H, H^{II}-3, H^{I}-6′), 3.44 – 3.34 (m, 2H, $H^{I,II}$-5), 3.31 (dd, *J* = 10.7, 3.8 Hz, 1H, H^{II}-6), 3.17 (dd, *J* = 10.7, 5.6 Hz, 1H, H^{II}-6′), 1.80 (s, 3H, Ac Hs), 1.19 (s, 9H, Piv Hs), 1.12 (s, 9H, Piv Hs). ^{13}C NMR (126 MHz, cdcl$_3$) δ 176.55, 176.26 (Piv C=Os), 169.70 (Ac C=O), 139.91, 138.63, 138.11, 137.89, 131.41, 128.94, 128.58, 128.48, 128.24, 128.09, 127.90, 127.79, 127.71, 127.51, 127.45, 127.36, 125.69 (arom Cs), 99.88 (C^{II}-1), 91.63 (C^{I}-1), 81.67 (C^{I}-3), 80.59 (C^{II}-3), 80.34 (C^{I}-5), 74.95 (C^{I}-4), 74.49, 73.83, 73.76, 73.72 (Ph$\underline{C}H_2$), 73.56 (C^{II}-5), 72.70 (C^{II}-2), 71.08 (C^{II}-4), 69.89 (C^{II}-6), 67.98 (C^{I}-2,6), 38.93, 38.88 (Piv 4˚ Cs), 27.37, 27.19 (Piv $\underline{C}H_3$), 20.89 (Ac $\underline{C}H_3$). ESI(+) HRMS calcd for $C_{58}H_{68}NaO_{14}S$ ([M+Na]$^+$): 1043.4222; found: 1043.4222.

Thiophenyl 4-*O*-acetyl-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranoside (3.13): The title compound was isolated from sulfoxide donor **3.12** (676 mg, 0.66 mmol) and acceptor **3.11** (381 mg, 0.40 mmol) with reagents DTBMP (487 mg, 2.38 mmol) and Tf_2O (1 M, 331 μL, 0.33 mmol) in CH_2Cl_2 (36 mL) as described with disaccharide **3.2**.

Disaccharide byproduct **3.2** (57 mg, 14% yield) was collected from flash chromatography at 15% EtOAc/ hexanes, followed by desired product as a white foam (454 mg, 61% yield).



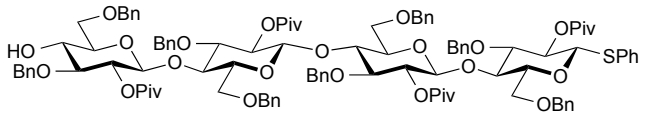
Chemical Formula: $C_{108}H_{128}O_{25}S$
Exact Mass: 1856.85
Molecular Weight: 1858.25

13

3.13: ^{1}H NMR (800 MHz, cdcl$_3$) δ 7.49 (d, *J* = 7.8 Hz, 2H, arom Hs), 7.37 – 7.06 (m, 43H, arom Hs), 5.17 (d, *J* = 11.9 Hz, 1H, PhC$\underline{H}_2$), 5.14 (d, *J* = 11.8 Hz, 1H, PhC$\underline{H}_2$), 5.04 (t, *J* = 9.6 Hz, 1H, H^{I}-2), 5.01 (dd, *J* = 9.4, 8.2 Hz, 1H, H^{IV}-2), 4.99 – 4.96 (m, 2H, H^{III}-2, PhC$\underline{H}_2$), 4.95 (t, *J* = 9.5 Hz, 1H, H^{IV}-4), 4.92 (t, *J* = 9.6 Hz, 1H, H^{II}-2), 4.71 (d, *J* = 12.0 Hz, 1H, PhC$\underline{H}_2$), 4.60 (d, *J* = 10.1 Hz, 1H, H^{I}-1), 4.55 (d, *J* = 11.2 Hz, 1H, PhC$\underline{H}_2$), 4.52 (d, *J* = 11.9 Hz, 1H, PhC$\underline{H}_2$), 4.48 (d, *J* = 11.2 Hz, 1H, PhC$\underline{H}_2$), 4.46 – 4.41 (m, 3H, PhC$\underline{H}_2$), 4.38 (d, *J* = 8.2 Hz, 1H, H^{III}-1), 4.36 (d, *J* = 8.0 Hz, 1H, H^{II}-1), 4.31 (d, *J* = 12.1 Hz, 1H, PhC$\underline{H}_2$), 4.27 (d, *J* = 8.2 Hz, 1H, H^{IV}-1), 4.22 (d, *J* = 11.8 Hz, 1H, PhC$\underline{H}_2$), 4.20 (d, *J* = 12.1 Hz, 1H, PhC$\underline{H}_2$), 4.08 – 4.04 (m, 2H, H^{I}-4, PhC$\underline{H}_2$), 4.03 (t, *J* = 9.6 Hz, 1H, H^{III}-4), 3.99 (d, *J* = 12.1 Hz, 1H, PhC$\underline{H}_2$), 3.96 (t, *J* = 9.0 Hz, 1H, H^{II}-4), 3.94 (d, *J* = 12.0 Hz, 1H, PhC$\underline{H}_2$), 3.78 (dd, *J* = 10.7, 3.3 Hz, 1H, H^{I}-6), 3.73 (d, *J* = 10.7 Hz, 1H, H^{I}-6'), 3.62 (t, *J* = 9.0 Hz, 1H, H^{I}-3), 3.57 (dd, *J* = 10.9, 2.8 Hz, 1H, H^{III}-6), 3.54 – 3.50 (m, 2H, H$^{III, IV}$-6'), 3.48 (dd, *J* = 10.7, 2.7 Hz, 1H, H^{IV}-6), 3.44 (t, *J* = 9.4 Hz, 1H, H^{IV}-3), 3.37 – 3.34 (m, 1H, H^{I}-5), 3.32 (t, *J* = 9.1 Hz, 1H, H^{III}-3), 3.31 – 3.28 (m,

1H, H^{IV}-5), 3.27 (dd, *J* = 10.6, 3.9 Hz, 1H, H^{II}-6), 3.22 (t, *J* = 9.2 Hz, 1H, H^{II}-3), 3.13 (bd, *J* = 9.8 Hz, 1H, H^{III}-5), 3.12 – 3.09 (m, 2H, H^{II}-5, H^{II}-6′), 1.79 (s, 3H, Ac Hs), 1.12 (s, 18H, Piv Hs), 1.08 (s, 9H, Piv Hs), 1.04 (s, 9H, Piv Hs). ^{13}C NMR (201 MHz, cdcl$_3$) δ 176.75, 176.72, 176.58 [2C] (Piv C=Os), 169.72 (Ac C=O), 139.36, 139.19, 139.02, 138.29, 138.21, 137.94, 137.82, 133.28, 132.66, 128.93, 128.77, 128.74, 128.53, 128.45, 128.39, 128.34, 128.33, 128.20, 128.18, 128.14, 128.06, 128.03, 127.89, 127.84, 127.74, 127.70, 127.54, 127.41, 127.20, 127.15, 127.01, 126.93, 126.79, 126.71 (arom Cs), 99.83 (C^{III}-1), 99.68 (C^{IV}-1), 99.46 (C^{II}-1), 86.66 (C^{I}-1), 82.76 (C^{I}-3), 81.42 (C^{III}-3), 80.79 (C^{II}-3), 80.52 (C^{IV}-3), 79.29 (C^{I}-5), 75.53 ($C^{II,III}$-4), 75.36 (C^{I}-4), 75.10 (C^{III}-5), 75.05 (C^{II}-5), 74.82, 74.63, 74.51, 73.86, 73.70 (Ph$\underline{C}$H$_2$), 73.55 ($C^{(V}$-5), 73.52, 73.45 (Ph$\underline{C}$H$_2$), 72.69 (C^{II-IV}-2), 71.30 (C^{IV}-4), 70.79 (C^{I}-2), 70.07 (C^{II}-6), 67.99 (C^{I}-6), 67.77 (C^{IV}-6), 67.40 (C^{III}-6), 38.86, 38.82, 38.77, 38.75 (Piv 4˚Cs), 27.39, 27.35, 27.24, 27.15 (Piv $\underline{C}$H$_3$), 20.88 (Ac $\underline{C}$H$_3$). ESI(+) HRMS calcd for $C_{108}H_{136}N_2O_{25}S$ ($[M+2(NH_4)]^{2+}$): 946.4571; found: 946.4569.

Thiophenyl 3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranoside (3.14): The title compound was isolated as a white foam (445 mg, 98% yield) from tetrasaccharide **3.13** (463 mg, 0.249 mmol) and methanolic NaOCH$_3$ (1.9 M, 33 μL, 0.06 mmol) in guanidine-buffered ethanol/ CH_2Cl_2 as described for compound **3.11**.



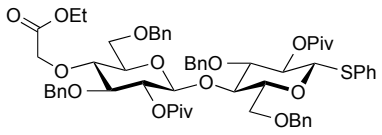
Chemical Formula: $C_{106}H_{126}O_{24}S$
Exact Mass: 1814.84
Molecular Weight: 1816.21

14

3.14: ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.41 (dd, *J* = 4.1, 3.3 Hz, 2H, SPh Hs), 7.33 – 6.96 (m, 43H, arom Hs), 5.13 (d, *J* = 11.9 Hz, 1H, PhC$\underline{H}_2$), 5.07 (d, *J* = 11.8 Hz, 1H, PhC$\underline{H}_2$), 4.98 (t, *J* = 9.6 Hz, 1H, H^{I}-2), 4.93 – 4.81 (m, 4H, H$^{II\text{-}IV}$-2, PhC$\underline{H}_2$), 4.71 (d, *J* = 11.4 Hz, 1H, PhC$\underline{H}_2$), 4.64 (d, *J* = 12.0 Hz, 1H, PhC$\underline{H}_2$), 4.54 – 4.49 (m, 2H, H^{I}-1, PhC$\underline{H}_2$), 4.44 (d, *J* = 11.9 Hz, 1H, PhC$\underline{H}_2$), 4.38 – 4.27 (m, 4H, H^{II}-1, PhC$\underline{H}_2$), 4.23 (apparent d, *J* = 10.2 Hz, 2H, H^{IV}-1, PhC$\underline{H}_2$), 4.21 – 4.11 (m, 4H, H^{III}-1, PhC$\underline{H}_2$), 4.01 – 3.93 (m, 2H, H^{I,II}-4), 3.91 – 3.86 (m, 2H, H^{III}-4, PhC$\underline{H}_2$), 3.84 (d, *J* = 12.0 Hz, 1H, PhC$\underline{H}_2$), 3.71 (bd, *J* = 10.3 Hz, 1H, H^{I}-6), 3.65 (d, *J* = 10.7 Hz, 1H, H^{I}-6'), 3.56 – 3.47 (m, 3H, H^{I}-3, H^{IV}-4, H^{II}-6), 3.46 – 3.41 (m, 3H, H^{II}-6', H^{III}-6,6'), 3.31 – 3.26 (m, 2H, H^{I}-5, H^{IV}-6), 3.23 (t, *J* = 9.2 Hz, 1H, H^{II}-3), 3.18 (t, *J* = 9.3 Hz, 1H, H^{IV}-3), 3.13 – 3.07 (m, 3H, H^{III}-3, H^{IV}-5,6'), 3.07 – 2.99 (m, 2H, H^{II,III}-5), 2.96 (s, 1H, RO$\underline{H}$), 1.04 (s, 9H, Piv Hs), 1.03 (s, 9H, Piv Hs), 0.99 (s, 9H, Piv Hs), 0.96 (s, 9H, Piv Hs); ^{13}C NMR (126 MHz, cdcl$_3$) δ 176.76, 176.67 (3C) [Piv C=Os], 139.34, 139.13 (2C), 138.43, 138.05, 137.77 (2C), 137.57 (4° arom Cs), 133.19 (C^{SPh} 4° C), 132.61, 128.87, 128.70, 128.68, 128.49, 128.38, 128.30, 128.26, 128.10, 128.05, 127.98, 127.83, 127.74, 127.70, 127.57, 127.26, 126.94, 126.87, 126.76, 126.66 (arom Cs), 99.76 (C^{II}-1), 99.62 (C^{III}-1), 99.34 (C^{IV}-1), 86.58 (C^{I}-1), 82.71 (C^{I}-3), 82.44 (C^{IV}-3), 81.37 (C^{II}-3), 80.64 (C^{III}-3), 79.23 (C^{I}-5), 75.51 (C^{II}-4), 75.27

($C^{I,III}$-4), 75.06 (C^{II}-5), 74.95 (C^{III}-5), 74.75, 74.42 (Ph<u>C</u>H_2), 74.33 (C^{IV}-4), 74.21, 73.79, 73.70, 73.44, 73.35 (Ph<u>C</u>H_2), 72.84 (C^{IV}-5), 72.74 (C^{II}-2), 72.61 ($C^{III,IV}$-2), 71.25 (C^{IV}-6), 70.73 (C^{I}-2), 67.93 (C^{I}-6), 67.68 (C^{III}-6), 67.35 (C^{II}-6), 38.79, 38.74, 38.69 (2C) [4° Piv Cs], 27.33, 27.27, 27.18, 27.10 (Piv <u>C</u>H_3); ESI(+) HRMS calcd for $C_{106}H_{134}N_2O_{24}S$ ($[M+2(NH_4)]^{2+}$): 925.4518; found: 925.4533.

Thiophenyl 4-*O*-(2-ethoxy-2-oxoethyl)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranoside (3.15): Sodium hydride (60% in mineral oli, 107 mg, 2.68 mmol) and tetrabutylammonium iodide (396 mg, 1.07 mmol) were stirred in anhydrous DMF (1.5 mL) at -20°C. Compound **3.11** (516 mg, 0.536 mmol) in anhydrous DMF (3.5 mL) was added drop-wise, and the solution was stirred under Ar for 1 hour, warming to 0°C. Ethyl bromoacetate (190 μL, 1.63 mmol) was added and the mixture stirred for 1 hour until the starting material had been consumed on TLC (30% EtOAc/ hexanes). The reaction was quenched with iced 10% $NaHCO_3$ (40 mL) and extracted with Et_2O (3 X 80 mL). The extracts were washed with water (4 X 50 mL) and brine (50 mL). The organic layer was dried over Na_2SO_4 and concentrated under vacuum. The crude was purified by flash chromatography (12 g SiO_2, 0—12% EtoAc/ hexanes), yielding pure **3.15** (487 mg, 87% yield) as a white foam.

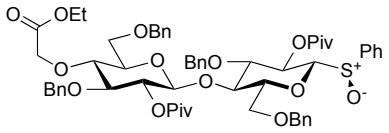


Chemical Formula: $C_{60}H_{72}O_{14}S$
Exact Mass: 1048.46
Molecular Weight: 1049.28

15

3.15: ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.52 – 7.47 (m, 2H, arom Hs), 7.37 – 7.10 (m, 23H, arom Hs), 5.10 – 4.99 (m, 3H, H^{I}-2, H^{II}-2, [C=O]C<u>H</u>), 4.75 (d, *J* = 11.1 Hz, 1H, PhC<u>H</u>$_2$), 4.68 (d, *J* = 11.9 Hz, 1H, PhC<u>H</u>$_2$), 4.65 – 4.46 (m, 5H, H^{I}-1, H^{II}-1, [C=O]C<u>H</u>', PhC<u>H</u>$_2$), 4.26 (s, 2H, PhC<u>H</u>$_2$), 4.20, 4.16 (ABq, *J* = 15.9 Hz, 2H, PhC<u>H</u>$_2$), 4.12 – 4.02 (m, 3H, H^{I}-4, OC<u>H</u>$_2$CH$_3$), 3.80 – 3.74 (m, 2H, H^{I}-6,6'), 3.68 (d, *J* = 11.2 Hz, 1H, H^{II}-6), 3.63 (t, *J* = 8.9 Hz, 1H, H^{I}-3), 3.60 – 3.50 (m, 3H, H^{II}-3,4,6'), 3.42 – 3.34 (m, 2H, H^{I}-5, H^{II}-5), 1.17 – 1.13 (m, 21H, OCH$_2$C<u>H</u>$_3$, PivC<u>H</u>$_3$). ^{13}C NMR (126 MHz, cdcl$_3$) δ 176.64, 176.57 (PivC=O), 169.71 (AcC=O), 138.89, 138.36, 138.03, 137.93 (4° arom Cs), 133.17 (arom S-<u>C</u>), 132.60, 128.84, 128.56, 128.37, 128.27, 128.00, 127.88, 127.83, 127.62, 127.50, 127.35, 127.20, 126.99 (arom Cs), 99.46 (C^{II}-1), 86.54 (C^{I}-1), 82.95 (C^{II}-3), 82.50 (C^{I}-3), 79.40 (C^{I}-5), 79.06 (C^{II}-4), 74.97 (C^{II}-5), 74.92 (Ph<u>C</u>H$_2$), 74.80 (C^{I}-4), 74.65 (Ac<u>C</u>H$_2$), 73.69 (C^{II}-2), 73.52, 73.29 (Ph<u>C</u>H$_2$), 70.85 (C^{I}-2), 69.96 (Ph<u>C</u>H$_2$), 68.55 (C^{II}-6), 68.24 (C^{I}-6), 60.75 (O<u>C</u>H$_2$CH$_3$), 38.83, 38.69 (4° Piv Cs), 27.31, 27.12 (Piv <u>C</u>H$_3$), 14.15 (OCH$_2$<u>C</u>H$_3$). ESI(+)-MS for $C_{60}H_{72}O_{14}SNa$ ([M+Na]$^+$): 1072.29; found: 1072.5.

Phenyl (4-*O*-(2-ethoxy-2-oxoethyl)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl)sulfoxide (3.16): The title compound was prepared as a mixture of sulfoxide diastereomers (646 mg, 98.2% yield) from compound **3.15** (648 mg, 0.618 mmol) and mCPBA (152 mg, 0.68 mmol) as described for compounds **3.12** and **3.7**.

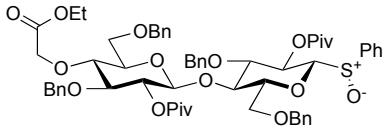


Chemical Formula: $C_{60}H_{72}O_{15}S$
Exact Mass: 1064.46
Molecular Weight: 1065.28

16a

3.16a (less polar sulfoxide): ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.82 (d, *J* = 8.3 Hz, 2H, S-arom Hs), 7.47 – 7.35 (m, 3H, arom Hs), 7.33 – 7.20 (m, 13H, arom Hs), 7.20 – 7.12 (m, 5H, arom Hs), 7.10 (bd, *J* = 7.7 Hz, 2H, arom Hs), 5.08 – 5.00 (m, 2H, H^{I}-2, AcC<u>H</u>), 4.97 (bt, *J* = 8.2 Hz, 1H, H^{II}-2), 4.71 (d, *J* = 11.1 Hz, 1H, PhC<u>H</u>$_2$), 4.59 (apparent t, *J* = 11.4 Hz, 2H, PhC<u>H</u>$_2$, AcC<u>H</u>'), 4.47 (apparent t, *J* = 8.6 Hz, 2H, H^{I}-1, H^{II}-1), 4.26, 4.23 (ABq, *J* = 12.1 Hz, 2H, PhC<u>H</u>$_2$), 4.21 – 4.11 (m, 4H, PhC<u>H</u>$_2$), 4.09 – 4.03 (m, 2H, OC<u>H</u>$_2$CH$_3$), 3.93 (t, *J* = 8.8 Hz, 1H, H^{I}-4), 3.79 – 3.71 (m, 2H, H^{I}-3,6), 3.67 (d, *J* = 12.1 Hz, 1H, H^{I}-6'), 3.60 – 3.49 (m, 4H, H^{II}-3,4,6,6'), 3.42 (bd, *J* = 9.5 Hz, 1H, H^{I}-5), 3.28 (bs, 1H, H^{II}-5), 1.22 – 1.14 (m, 12H, OCH$_2$C<u>H</u>$_3$, Piv-Hs), 1.12 (s, 9H, Piv-Hs). ^{13}C NMR (101 MHz, cdcl$_3$) δ 177.04, 176.65 (PivC=O), 169.74 (AcC=O), 138.55, 138.32, 138.00, 137.77, 131.72 (4° arom Cs), 128.71, 128.48, 128.41, 128.31, 128.15, 127.85, 127.74, 127.67, 127.50, 127.41, 127.35, 127.26, 127.21, 126.58 (arom Cs),

99.61 (C^{II}-1), 94.02 (C^{I}-1), 82.95 (C^{II}-3), 81.73 (C^{I}-3), 79.86 (C^{I}-5), 78.99 (C^{II}-4), 75.02 (Ph<u>C</u>H_2), 74.86 (C^{II}-5, Ac<u>C</u>H_2), 74.22 (C^{I}-4), 73.73 (Ph<u>C</u>H_2), 73.40 (C^{II}-2), 73.23 (Ph<u>C</u>H_2), 70.03 (Ph<u>C</u>H_2), 68.41 (C^{I}-2, C^{II}-6), 67.74 (C^{I}-6), 60.80 (O<u>C</u>H_2CH_3), 38.89, 38.86 (4° Piv-Cs), 27.29, 27.08 (Piv <u>C</u>H_3), 14.16 (OCH_2<u>C</u>H_3).



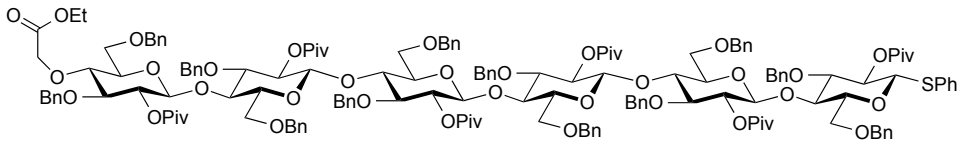
Chemical Formula: $C_{60}H_{72}O_{15}S$
Exact Mass: 1064.46
Molecular Weight: 1065.28

16b

3.16b (more polar sulfoxide): ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.65 – 7.60 (m, 2H, arom Hs), 7.47 – 7.40 (m, 3H, arom Hs), 7.38 – 7.13 (m, 20H, arom Hs), 5.41 (t, *J* = 8.8 Hz, 1H, H^{I}-2), 5.03 – 4.94 (m, 2H, H^{II}-2, AcC<u>H</u>), 4.73 (d, *J* = 11.1 Hz, 1H, PhC<u>H</u>$_2$), 4.64 – 4.54 (m, 3H, PhC<u>H</u>$_2$, AcC<u>H</u>′), 4.49 (d, *J* = 8.0 Hz, 1H, H^{II}-1), 4.40 (d, *J* = 11.9 Hz, 1H, PhC<u>H</u>$_2$), 4.28, 4.26 (ABq, *J* = 12.1 Hz, 2H, PhC<u>H</u>$_2$), 4.22 – 4.03 (m, 6H, H^{I}-1, H^{I}-4, PhC<u>H</u>$_2$, OC<u>H</u>$_2CH_3$), 3.73 (t, *J* = 8.2 Hz, 1H, H^{I}-3), 3.69 – 3.60 (m, 2H, H^{I}-6, H^{II}-6), 3.60 – 3.52 (m, 4H, H^{I}-6′, H^{II}-3,4,6′), 3.38 – 3.34 (m, 2H, H^{I}-5, H^{II}-5), 1.18 (t, *J* = 7.1 Hz, 3H, OCH_2C<u>H</u>$_3$), 1.15 (s, 9H, Piv-Hs), 1.11 (s, 9H, Piv-Hs). ^{13}C NMR (126 MHz, cdcl$_3$) δ 176.57, 176.13 (PivC=O), 169.66 (AcC=O), 139.78, 138.60, 138.21, 137.94, 137.90 (4° arom Cs), 131.27, 128.79, 128.45, 128.33, 128.28, 128.03, 127.90, 127.78, 127.59, 127.51, 127.39, 127.18, 127.06, 125.59 (arom Cs), 99.71 (C^{II}-1), 91.53 (C^{I}-1), 82.90 (C^{II}-3), 81.60 (C^{I}-3), 80.18 (C^{II}-5), 78.93 (C^{II}-4), 74.90 (C^{I}-5, Ph<u>C</u>H_2), 74.38 (C^{I}-4), 74.08 (Ac<u>C</u>H_2), 73.67 (Ph<u>C</u>H_2), 73.36 (C^{II}-2), 73.27 (Ph<u>C</u>H_2), 69.94 (Ph<u>C</u>H_2), 68.43

(C^{II}-6), 68.02 (C^{I}-6), 67.72 (C^{I}-2), 60.74 ($O\underline{C}H_2CH_3$), 38.76, 38.74 (4° Piv-Cs), 27.24, 27.01 ($Piv\underline{C}H_3$), 14.11 ($OCH_2\underline{C}H_3$).

Thiophenyl 4-*O*-(2-ethoxy-2-oxoethyl)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranoside (3.17): The title compound was prepared from sulfoxide donor **3.16** (646 mg, 0.607 mmol) and acceptor **3.14** (722 mg, 0.40 mmol) with reagents DTBMP (490 mg, 2.39 mmol) and Tf_2O (1 M, 430 μL, 0.43 mmol) in CH_2Cl_2 (67 mL) as described for disaccharide **3.2** and tetrasaccharide **3.13**. The reaction was complete by TLC (30% EtOAc/ hexanes) upon warming to -20°C and stirring for 4 hrs. The reaction concentrate was dissolved in $CHCl_3$ (10 mL) and EtOH (10 mL) while hot, and the crude product was precipitated in excess EtOH (30 mL). The crude solid was dissolved in hot EtOH (20 mL), and upon cooling to -20°C, the product precipitated as a white, waxy solid (561 mg, 51.2% yield).



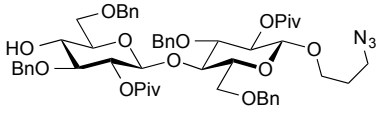
Chemical Formula: $C_{160}H_{192}O_{38}S$
Exact Mass: 2753.28
Molecular Weight: 2755.32

17

3.17: ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.53 – 7.41 (m, 2H, arom Hs), 7.40 – 6.95 (m, 63H, arom Hs), 5.23 – 5.01 (m, 5H, H^{I}-2, PhC$\underline{H}_2$), 5.00 – 4.87 (m, 4H, H^{II-VI}-2, AcC$\underline{H}$), 4.74 – 4.68 (m, 2H, PhC$\underline{H}_2$), 4.62 – 4.55 (m, 2H, H^{I}-1, PhC$\underline{H}_2$), 4.51 (d, *J* = 11.8 Hz, 1H, PhC$\underline{H}_2$), 4.48 – 4.11 (m, 19H, H^{II-VI}-1, PhC$\underline{H}_2$, AcC$\underline{H}$'), 4.10 – 3.91 (m, 10H, H^{I-V}-4, PhC$\underline{H}_2$, OC$\underline{H}_2$CH$_3$), 3.90 – 3.84 (m, 2H, PhC$\underline{H}_2$), 3.82 – 3.69 (m, 3H, H^{I}-6,6', H^{IV}-6), 3.65 – 3.58 (m, 2H, H^{I}-3, H^{VI}-6'), 3.58 – 3.41 (m, 10H, H^{VI}-3,4, H^{II-V}-6,6'), 3.39 – 3.02 (m, 10H, H^{II-V}-3, H^{I-6}-5), 1.39 – 0.76 (m, 57H, PivHs, OCH$_2$C$\underline{H}_3$). ^{13}C NMR (126 MHz, cdcl$_3$) δ 176.65 (2C), 176.60 (3C), 176.56 (PivC=O), 169.69 (AcC=O), 139.34, 139.31, 139.26, 139.12, 139.08, 138.39, 138.03, 138.00, 137.91, 137.78, 137.71(2C) [4° arom Cs], 133.19 (C^{SPh} 4° arom), 132.53, 128.80, 128.61, 128.55, 128.43, 128.32, 128.23, 128.18, 128.06, 127.91, 127.77, 127.72, 127.63, 127.55, 127.44, 127.33, 127.23, 127.14, 127.08, 127.03, 126.90, 126.85, 126.81, 126.65, 126.57, 126.53 (arom Cs), 99.73, 99.58, 99.53, 99.44, 99.34 (C^{II-VI}-1), 86.56 (C^{I}-1), 82.85, 82.65, 81.31, 81.22, 81.19, 81.08 (C^{I-VI}-3), 79.19 (C^{VI}-4), 79.09 (C^{I}-5), 75.51, 75.46, 75.40, 75.24, 75.02, 74.83, 74.78, 74.71 (C^{II-VI}-5, Ph$\underline{C}$H$_2$), 74.33, 73.73, 73.64, 73.40, 73.34, 73.22, 72.50 (Ph$\underline{C}$H$_2$, Ac$\underline{C}$H$_2$), 70.69 (C^{I}-2), 69.92 (Ph$\underline{C}$H$_2$), 68.67, 67.90, 67.77, 67.67, 67.40, 67.31 (C^{I-VI}-6), 60.70 (O$\underline{C}$H$_2$CH$_3$), 38.77 (s, 2C), 38.62 (4C, 4° Piv-Cs), 27.27, 27.11 (4C), 27.03 (Piv$\underline{C}$H$_3$), 14.10 (OCH$_2$$\underline{C}H_3$). ESI(+)-MS calculated for C$_{160}$H$_{192}$O$_{38}$SNa$_2$ ([M+2Na]$^{2+}$): 1400.6; found 1399.6.

3-azidopropyl 3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranoside (3.18): Compound **3.11** (575 mg, 0.598 mmol) was

combined with 4Å molecular sieves (~500 mg) in a flame-dried 3-neck flask under Ar. The mixture was diluted with anhydrous CH_2Cl_2 (10 mL) and 3-azidopropanol (300 mg, 2.99 mmol; *vide infra*) in CH_2Cl_2 (5 mL) was added. The mixture stirred at rt for 45 min and was then cooled to -20°C for 15 min (iced NaCl). Solid AgOTf was azeotropically dried from toluene and added (77 mg, 0.3 mmol) along with N-iodosuccinimide (216 mg, 0.96 mmol) to the reaction mixture, causing the mixture to become pale yellow/ orange. The reaction warmed to 15°C over 7 hrs, at which point the mixture had become dark fuchsia. The reaction was diluted with CH_2Cl_2 (150 mL) and filtered through celite. The filtrate was washed with 5% $NaHCO_3$ (50 mL), 10% $NaHSO_3$ (50 mL), water (2 X 50 mL), and brine (50 mL) then dried over Na_2SO_4. The crude material was loaded onto SiO_2 (1 g) and purified by flash chromatography (2 X 24 g SiO_2, 0—15% EtOAc/ hexanes, 35 mL/ min). The title compound was obtained as a colorless gum (514 mg, 90% yield).



Chemical Formula: $C_{53}H_{67}N_3O_{13}$
Exact Mass: 953.47
Molecular Weight: 954.13

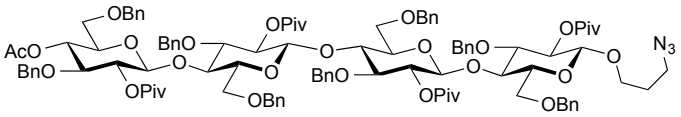
18

3.18: ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.39 – 7.14 (m, 20H, arom Hs), 5.01 – 4.89 (m, 3H, H^{I}-2, H^{II}-2, PhC$\underline{H}_2$), 4.83 – 4.75 (m, 2H, PhC$\underline{H}_2$), 4.62 (d, *J* = 11.4 Hz, 1H, PhC$\underline{H}_2$), 4.54 (d, *J* = 11.2 Hz, 1H, PhC$\underline{H}_2$), 4.49 – 4.40 (m, 2H, H^{II}-1, PhC$\underline{H}_2$), 4.38 – 4.29 (m, 3H, H^{I}-1, PhC$\underline{H}_2$), 4.04 (t, *J* = 9.3 Hz, 1H, H^{I}-4), 3.89 (dt, *J* = 10.3, 5.8 Hz, 1H, OC$\underline{H}_2$CH$_2$CH$_2$N$_3$), 3.77 (dd, *J* =

10.9, 3.4 Hz, 1H, H^{I}-6), 3.72 – 3.67 (m, 1H, H^{I}-6'), 3.67 – 3.63 (m, 1H, H^{II}-4), 3.58 (t, *J* = 9.2 Hz, 1H, H^{I}-3), 3.55 – 3.46 (m, 2H, H^{II}-6, $OCH_2CH_2CH_2N_3$), 3.39 – 3.24 (m, 6H, H^{I}-5, H^{II}-3,5,6', $OCH_2CH_2CH_2N_3$), 2.98 (s, 1H, OH), 1.89 – 1.74 (m, 2H, $OCH_2CH_2CH_2N_3$), 1.17 – 1.12 (m, 18H, Piv-Hs). ^{13}C NMR (126 MHz, $cdcl_3$) δ 176.69 (2C, PivC=O), 138.88, 138.37, 137.80, 137.51 (4° arom Cs), 128.61, 128.48, 128.35, 128.22, 128.00, 127.96, 127.82, 127.67, 127.56, 127.27, 127.21, 127.02 (arom Cs), 101.31 (C^{I}-1), 99.41 (C^{II}-1), 82.44 (C^{II}-3), 80.74 (C^{I}-3), 75.13 (C^{I}-5), 75.08 (C^{I}-4), 74.36, 74.29 ($PhCH_2$), 74.03 (C^{II}-4), 73.72 (2C, $PhCH_2$), 72.98 (C^{II}-5), 72.75 (C^{II}-2), 72.15 (C^{I}-2), 67.82 (C^{II}-6), 65.96 (C^{I}-6), 48.09 ($OCH_2CH_2CH_2N_3$), 38.77, 38.73 (4° Piv-Cs), 29.15 ($OCH_2CH_2CH_2N_3$), 27.27, 27.11 ($PivCH_3$). ESI(+)-MS calculated for $C_{53}H_{67}N_3O_{13}Na$ ($[M+Na]^+$): 976.47; found: 976.5.

3-azidopropyl 4-*O*-acetyl-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranoside (3.19): 4ÅMS (500 mg) and a stir bar were flame dried under house vacuum in a three neck rb flask then cooled to rt under Ar. A solution of compounds **3.2** (237 mg, 0.236 mmol) and **3.18** (150 mg, 0.157 mmol) in anhydrous CH_2Cl_2 (5 mL) was added via cannula, and the mixture was stirred for 45 min. at rt then cooled to -20 °C (ice/ NaCl). After 10 min, solid NIS (71 mg, 0.314 mmol) and AgOTf (40 mg, 0.157) were added. The mixture slowly warmed to 22 °C over 12 hrs, at which point the mixture became dark red and

compound **2** was no longer observed via TLC (25% EtOAc/ hexanes). The mixture was diluted with EtOAc (200 mL) and filtered through celite. The filtrate was washed with 5% $NaHCO_3$ (50 mL), 10% $NaHSO_3$ (50 mL), water (2 X 50 mL), and brine (50 mL) then dried over anhydrous Na_2SO_4. The crude product was concentrated under vacuum onto SiO_2 (1.2 g) and purified by flash chromatography over SiO_2 (2 X 12 g; 0 – 18% EtOAc/ hexanes to remove impurities then 20% EtOAc to elute product). Tetrasaccharide **3.19** (171 mg, 59% yield) was obtained as a white solid.



Chemical Formula: $C_{105}H_{129}N_3O_{26}$
Exact Mass: 1847.89
Molecular Weight: 1849.18

19

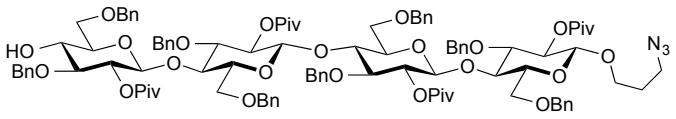
3.19: ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.37 – 7.07 (m, 40H, arom Hs), 5.18 (d, *J* = 12.0 Hz, 1H, PhC$\underline{H}_2$), 5.13 (d, *J* = 11.9 Hz, 1H, PhC$\underline{H}_2$), 5.04 – 4.92 (m, 5H, H$^{I\text{-}IV}$-2, H^{IV}-4), 4.78 (d, *J* = 12.2 Hz, 1H, PhC$\underline{H}_2$), 4.58 – 4.39 (m, 7H, PhC$\underline{H}_2$), 4.38 – 4.31 (m, 4H, H$^{II\text{-}IV}$-1, PhC$\underline{H}_2$), 4.27 (d, *J* = 8.2 Hz, 1H, H^{I}-1), 4.24 – 4.18 (m, 2H, PhC$\underline{H}_2$), 4.09 – 4.00 (m, 4H, H$^{II\text{-}III}$-4, PhC$\underline{H}_2$), 3.99 – 3.87 (m, 3H, H^{I}-4, PhC$\underline{H}_2$, OC$\underline{H}_2$CH$_2$CH$_2$N$_3$), 3.80 (dd, *J* = 10.7, 3.1 Hz, 1H, H^{IV}-6), 3.67 (d, *J* = 9.3 Hz, 1H, H^{IV}-6'), 3.63 – 3.47 (m, 6H, H^{II}-6,6', H^{III}-3,6,6', OC$\underline{H}_2$CH$_2$CH$_2$N$_3$), 3.44 (t, *J* = 9.4 Hz, 1H, H^{IV}-3), 3.38 – 3.20 (m, 7H, H$^{I\text{-}II}$-3, H$^{III\text{-}IV}$-5, H^{I}-6, OCH$_2$CH$_2$C$\underline{H}_2$N$_3$), 3.16 – 3.08 (m, 3H, H$^{I\text{-}II}$-5, H^{I}-6'), 1.86 – 1.76 (m, 5H, OCH$_2$C$\underline{H}_2$CH$_2$N$_3$, AcC$\underline{H}_3$), 1.12 (s, 9H, Piv-Hs), 1.10 – 1.07 (m, 18H, Piv-Hs), 1.04 (s, 9H, Piv-Hs). ^{13}C NMR (126 MHz, cdcl$_3$) δ 176.85, 176.72 (2C), 176.56 (PivC=O), 169.70 (AcC=O), 139.40, 139.23, 139.03, 138.31,

138.23, 137.95, 137.85, 137.67 (4˚ arom Cs), 128.88, 128.75, 128.52, 128.45, 128.37, 128.14, 128.04, 127.84, 127.74, 127.70, 127.53, 127.42, 127.17, 126.95, 126.79, 126.71 (arom Cs), 101.43 (C^{IV}-1), 99.73, (2C), 99.47 (C^{I-III}-1), 81.37, 81.27, 80.82 (C^{I-III}-3), 80.54 (C^{IV}-3), 75.55, 75.46 (C^{I-III}-4), 75.08 (C^{I-III}-5), 74.51 (C^{IV}-4, Ph$\underline{C}H_2$), 73.93, 73.70 (Ph$\underline{C}H_2$), 73.54 (C^{IV}-5, Ph$\underline{C}H_2$), 72.70 (2C), 72.19, 71.30 (C^{I-IV}-2), 70.08 (C^{I}-6), 67.78, 67.71, 67.63, 67.46 (C^{II-IV}-6), 66.07 ($O\underline{C}H_2CH_2CH_2N_3$), 48.22 ($OCH_2CH_2\underline{C}H_2N_3$), 38.87, 38.84, 38.80, 38.76 (4˚ Piv-Cs), 29.28 ($OCH_2\underline{C}H_2CH_2N_3$), 27.39 (2C), 27.24, 27.17 (Piv$\underline{C}H_3$), 20.88 (Ac$\underline{C}H_3$).

3-azidopropyl 3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranoside (3.20):

Compound **3.19** (171 mg, 0.092 mmol) was dissolved in 4:1 EtOH/ CH_2Cl_2 (1.5 mL) and stirred openly. A solution of basic guanidine (0.46 mL, 0.092 mmol; 200 mM in 7:1 EtOH/ CH_2Cl_2) was added to the starting compound. A solution of sodium methoxide in methanol (1.9 M; 12 μL, 0.023 mmol) was added and the mixture stirred capped for 20 hrs, at which point the reaction was complete by TLC (25% EtOAc/ hexanes). The reaction was quenched with 10% aqueous AcOH (~1 mL) and neutralized with 5% $NaHOC_3$ (~5 mL). The product was extracted into CH_2Cl_2 (200 mL) and washed with water (2 X 50 mL) and brine (50 mL) then dried over anhydrous Na_2SO_4. The crude was dried to a foam under reduced pressure, and the foam was dissolved in hot EtOH (2

mL). Upon cooling to 4 °C and storage for 3 days, white powdery crystals formed and were collected by filtration. The cake was washed with 80% EtOH/ water and dried on the lyophilizer for 4 days to yield pure **3.20** (142 mg, 85% yield).



Chemical Formula: $C_{103}H_{127}N_3O_{25}$
Exact Mass: 1805.88
Molecular Weight: 1807.15

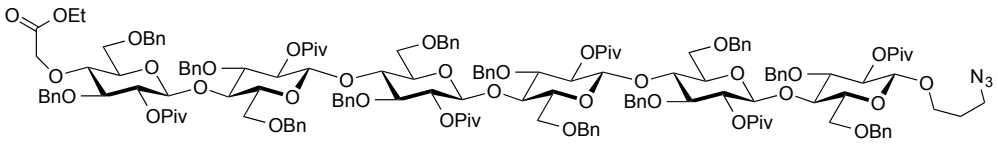
20

3.20: ^{1}H NMR (500 MHz, cdcl$_3$) δ 7.46 – 7.08 (m, 40H, arom Hs), 5.24 (d, *J* = 11.8 Hz, 1H, PhC$\underline{H}_2$), 5.17 (d, *J* = 11.9 Hz, 1H, , PhC$\underline{H}_2$), 5.09 – 4.92 (m, 5H, H$^{I\text{-}IV}$-2, PhC$\underline{H}_2$), 4.85 – 4.79 (m, 2H, PhC$\underline{H}_2$), 4.63 (d, *J* = 11.4 Hz, 1H, PhC$\underline{H}_2$), 4.56 (d, *J* = 12.0 Hz, 1H, PhC$\underline{H}_2$), 4.48 – 4.42 (m, 3H, PhC$\underline{H}_2$), 4.42 – 4.22 (m, 8H, H$^{I\text{-}IV}$-1, PhC$\underline{H}_2$), 4.16 – 3.90 (m, 6H, H$^{I\text{-}III}$-4, OC$\underline{H}_2$CH$_2$CH$_2$N$_3$, PhC$\underline{H}_2$), 3.84 (d, *J* = 10.6 Hz, 1H, H^{I}-6), 3.71 (d, *J* = 10.6 Hz, 1H, H^{I}-6′), 3.67 – 3.50 (m, 7H, H^{I}-3, H^{IV}-4, H$^{II\text{-}III}$-6,6′, OC$\underline{H}_2$CH$_2$CH$_2$N$_3$), 3.44 – 3.20 (m, 9H, H$^{II\text{-}IV}$-3, H$^{I\text{-}II}$-5, H^{IV}-6,6′, OCH$_2$CH$_2$C$\underline{H}_2$N$_3$), 3.16 (apparent dd, *J* = 16.1, 10.0 Hz, 2H, H$^{III\text{-}IV}$-5), 3.04 (s, 1H, O$\underline{H}$), 1.92 – 1.78 (m, 2H, OCH$_2$C$\underline{H}_2$CH$_2$N$_3$), 1.15 (s, 9H, Piv-Hs), 1.12 (s, 18H, Piv-Hs), 1.08 (s, 9H, Piv-Hs). ^{13}C NMR (126 MHz, cdcl$_3$) δ 176.84, 176.82, 176.75, 176.68 (PivC=O), 139.43, 139.22, 139.18, 138.48, 138.11, 137.83, 137.64, 137.61 (4˚ arom Cs), 128.86, 128.73, 128.54, 128.51, 128.43, 128.35, 128.15, 128.10, 128.03, 127.89, 127.75, 127.62, 127.32, 127.15, 126.96, 126.92, 126.80, 126.69 (arom Cs), 101.42 (C^{I}-1), 99.72 (2C, C$^{II\text{-}III}$-1), 99.39 (C^{IV}-1), 82.49, 81.37, 81.26, 80.69 (C$^{I\text{-}IV}$-3), 75.59, 75.43, 75.29 (C$^{I\text{-}III}$-4), 75.10, 74.99 (C$^{I\text{-}III}$-5), 74.51, 74.47, 74.38, 74.35, 74.31(C^{IV}-4, Ph$\underline{C}$H$_2$), 73.91, 73.76, 73.52, 73.40 (Ph$\underline{C}$H$_2$), 72.83, 72.79,

72.65, 72.17 (C^{I-IV}-2, C^{IV}-5), 71.34 (C^{IV}-6), 67.73, 67.67 , 67.42 (C^{I-III}-6), 66.05 ($O\underline{C}H_2CH_2CH_2N_3$), 48.20 ($OCH_2CH_2\underline{C}H_2N_3$), 38.85, 38.81, 38.78, 38.74 (4° Piv-Cs), 29.26 ($OCH_2\underline{C}H_2CH_2N_3$), 27.37, 27.33, 27.22, 27.15 ($Piv\underline{C}H_3$).

3-azidopropyl 4-*O*-(2-ethoxy-2-oxoethyl)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranosyl-(1→4)-3,6-*O*-benzyl-2-*O*-pivaloyl-*β*-D-glucopyranoside (3.21): *Method 1.* Compound **3.17** (45.3 mg, 0.016 mmol) was stirred with flame-dried 4Å molecular sieves (200 mg) in CH_2Cl_2 (1.5 mL) under Ar in a flame-dried flask. A solution of 3-azidopropanol in CH_2Cl_2 (1 M, 0.16 mL, 0.16 mmol; *vide infra*) was added and the mixture stirred at rt for 1 hr and then cooled to 0°C. A freshly-prepared solution of iodine in CH_2Cl_2 (0.1 M, 0.32 mL, 0.032 mmol) was added, causing solution to become dark red. A freshly-prepared solution of silver trifluoromethanesulfonate (azeotropically dried with toluene) in toluene (0.1 M, 0.32 mL, 0.032 mmol) was added drop-wise, causing the red color to dissipate, and the mixture was warmed to rt. After 1 hr the reaction as quenched with triethylamine (100 μL) and filtered over celite. The filtrate was concentrated onto SiO_2 (200 mg) and subsequently purified by flash chromatography (4 g SiO_2, 18 mL/ min, 0→12% →16% EtoAc/ hexanes) to yield title compound as a white foam (15 mg, 34% yield).

Method 2: The title compound could also be prepared via the glycosylation procedure for the formation of **3.19** from **3.2** and **3.18** described above using donor **3.15** and acceptor **3.20**.



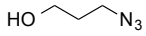
Chemical Formula: $C_{157}H_{193}N_3O_{39}$
Exact Mass: 2744.32
Molecular Weight: 2746.25

21

3.21: ^{1}H NMR (800 MHz, cdcl$_3$) δ 7.37 – 7.05 (m, 60H, arom Hs), 5.22 – 5.17 (m, 2H, PhC$\underline{H}_2$), 5.15 (d, *J* = 11.8 Hz, 1H, PhC$\underline{H}_2$), 5.12 (d, *J* = 11.8 Hz, 1H, PhC$\underline{H}_2$), 5.09 (d, *J* = 11.5 Hz, 1H, PhC$\underline{H}_2$), 5.00 (dd, *J* = 9.4, 8.1 Hz, 1H, H^{I}-2), 4.97 – 4.91 (m, 5H, H^{II-VI}-2), 4.76 (d, *J* = 12.1 Hz, 1H, PhC$\underline{H}_2$), 4.71 (d, *J* = 11.1 Hz, 1H, PhC$\underline{H}_2$), 4.58 (d, *J* = 11.1 Hz, 1H, PhC$\underline{H}_2$), 4.52 (d, *J* = 11.8 Hz, 1H, PhC$\underline{H}_2$), 4.42 – 4.31 (m, 8H), 4.26 (d, *J* = 8.2 Hz, 1H), 4.24 – 4.20 (m, 4H) [H^{I-VI}-1, PhC$\underline{H}_2$], 4.19 – 4.15 (m, 4H, PhC$\underline{H}_2$, AcC$\underline{H}$), 4.11 (d, *J* = 16.0 Hz, 1H, AcC$\underline{H}$'), 4.09 – 4.04 (m, 3H, H^{I}-4, OC$\underline{H}_2$CH$_3$), 4.04 – 3.95 (m, 6H, H^{II-V}-4, PhC$\underline{H}_2$), 3.94 (d, *J* = 12.1 Hz, 1H, PhC$\underline{H}_2$), 3.91 – 3.84 (m, 4H, PhC$\underline{H}_2$, OC$\underline{H}_2$CH$_2$CH$_2$N$_3$), 3.78 (dd, *J* = 10.7, 3.2 Hz, 1H, H^{I}-6), 3.66 (d, *J* = 9.2 Hz, 1H, H^{I}-6'), 3.63 – 3.43 (m, 14H, H^{I,VI}-3, H^{VI}-4, H^{II-VI}-6,6', OC$\underline{H}_2$CH$_2$CH$_2$N$_3$), 3.38 – 3.31 (m, 2H, OCH$_2$CH$_2$C$\underline{H}_2$N$_3$), 3.31 – 3.26 (m, 2H, H^{I,VI}-5), 3.24 (t, *J* = 9.2 Hz, 1H), 3.20 (t, *J* = 9.2 Hz, 2H), 3.16 (t, *J* = 9.3 Hz, 1H)[H^{II-V}-3], 3.13 (bd, *J* = 9.4 Hz, 1H), 3.10 – 3.03 (m, 3H)[H^{II-V}-5], 1.87 – 1.75 (m, 2H, OCH$_2$C$\underline{H}_2$CH$_2$N$_3$), 1.17 (t, *J* = 7.1 Hz, 3H, OCH$_2$C$\underline{H}_3$), 1.09 (s, 9H, Piv-Hs), 1.08 (s, 9H, Piv-Hs), 1.02 (s, 9H, Piv-Hs), 1.01 –

0.99 (m, 27H, Piv-Hs). ^{13}C NMR (201 MHz, cdcl$_3$) δ 176.89, 176.82, 176.80, 176.77, 176.71, 176.70 (PivC=O), 169.84 (AcC=O), 139.47 (2C), 139.39, 139.24 (2C), 138.52, 138.15, 138.11, 137.90, 137.83, 137.82, 137.65 (4° arom Cs), 128.90, 128.77, 128.75, 128.70, 128.56, 128.46, 128.42, 128.39, 128.35, 128.31, 128.14, 128.08, 128.05, 127.77, 127.69, 127.58, 127.37, 127.21, 127.17, 126.97, 126.93, 126.82, 126.81, 126.78, 126.74, 126.67, 126.64 (arom Cs), 101.45 (C^{I}-1), 99.76, 99.71, 99.66 (2C), 99.47 (C^{II-VI}-1), 82.97, 81.38, 81.35, 81.31, 81.28, 81.19 (C^{I-VI}-3), 79.20 (C^{VI}-4), 75.67, 75.58, 75.53, 75.48 (C^{II-V}-4), 75.24 (C^{I}-4), 75.14, 75.12, 75.10, 75.00, 74.94, 74.92, 74.88 (C^{I-VI}-5, Ph$\underline{C}H_2$), 74.56, 74.46, 73.94, 73.55, 73.52, 73.48, 73.45, 73.34 (Ph$\underline{C}H_2$), 72.62, 72.61, 72.19, 72.17 (C^{I-VI}-2), 70.05 (Ac$\underline{C}H_2$), 68.79, 67.87, 67.67, 67.53, 67.48, 67.43 (C^{I-VI}-6), 66.10 (O$\underline{C}H_2CH_2CH_2N_3$), 60.85 (O$\underline{C}H_2CH_3$), 48.23 ($OCH_2CH_2\underline{C}H_2N_3$), 38.90, 38.81, 38.76, 38.75 (3C)[4° Piv-Cs], 29.28 ($OCH_2\underline{C}H_2CH_2N_3$), 27.41, 27.29, 27.25 (2C), 27.23, 27.18 (Piv$\underline{C}H_3$), 14.23 ($OCH_2\underline{C}H_3$).

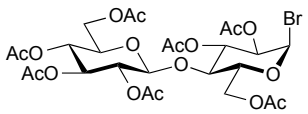
3-azido-1-propanol: 1-chloropropanol (3 g, 31.75 mmol), sodium azide (5.16 g, 79.4 mmol), and water (15 mL) were combined and stirred openly at 90 °C for 23 hours. Upon cooling to rt, the mixture was diluted with water (50 mL) and extracted with CH_2Cl_2 (3 X 50 mL). The extracts were washed with brine (50 mL) and dried over anhydrous Na_2SO_4. The solvents were removed under reduced pressure and the product obtained as a pale yellow oil (2.5 g, 78% yield).



Chemical Formula: $C_3H_7N_3O$
Exact Mass: 101.06
Molecular Weight: 101.11

IR (neat) ν 3334.94 (O-H), 2044.27, 2879.89 (C-H), 2090.33 ($N=N^+=N^-$), 1043 (C-O); ^1H NMR (400 MHz, cdcl3) δ 3.77 – 3.70 (m, 2H), 3.45 (t, *J* = 6.6 Hz, 2H), 2.46 – 2.39 (m, 1H), 1.83 (quint, 2H); ^{13}C NMR (75 MHz, cdcl3) δ 59.76, 48.47, 31.47.

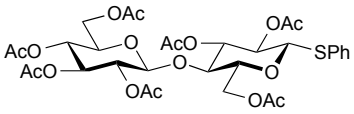
acetobromocellobiose:[292, 294] D-cellobiose (mostly β; 10 g, 27.8 mmol) was stirred in acetic anhydride (50 mL, 530 mmol) openly over an ice bath. A solution of HBr in acetic acid (33 wt%; 10 mL) was added drop-wise with an addition funnel. The mixture stirred for 15 min cold and then warmed to rt for 45 min. The slurry was supplemented with more HBr solution (50 mL) by drop-wise addition over 3 hrs at rt. The mixture stirred at rt over night. After 12 hrs, the intermediate sugar acetate (R_f = 0.1) had converted fully to the bromo sugar (R_f = 0.3) as analyzed by TLC (1:1 EtOAc/ hexanes). The reaction solvents were removed by co-distillation with toluene (3 X 50 mL) under reduced pressure. The resulting off-white solid was stirred in dry Et_2O (150 mL), and heated to a gentle reflux for 1 hr. The mixture was cooled on ice, filtered, and washed with excess Et_2O to yield desired product as a white, fluffy solid (19.6 g, 100% yield).



Chemical Formula: $C_{26}H_{35}BrO_{17}$
Exact Mass: 698.11
Molecular Weight: 699.45

^{1}H NMR (500 MHz, cdcl$_3$) δ 6.54 (d, *J* = 4.0 Hz, 1H), 5.54 (t, *J* = 9.7 Hz, 1H), 5.16 (t, *J* = 9.3 Hz, 1H), 5.09 (t, *J* = 9.7 Hz, 1H), 4.95 (dd, *J* = 9.2, 8.1 Hz, 1H), 4.77 (dd, *J* = 10.0, 4.1 Hz, 1H), 4.57 – 4.52 (m, 2H), 4.39 (dd, *J* = 12.5, 4.3 Hz, 1H), 4.23 – 4.15 (m, 2H), 4.05 (dd, *J* = 12.5, 2.1 Hz, 1H), 3.85 (t, *J* = 9.6 Hz, 1H), 3.68 (X of ABX, apparent ddd, *J* = 9.8, 4.2, 2.2 Hz, 1H), 2.15 (s, 3H), 2.10 (s, 6H), 2.06 (s, 6H), 2.02 (s, 3H), 2.00 (s, 3H). ^{13}C NMR (126 MHz, cdcl$_3$) δ 170.60, 170.37, 170.21, 170.09, 169.39 (2C), 169.09, 100.67, 86.50, 75.32, 73.07, 73.00, 72.10, 71.63, 70.81, 69.46, 67.76, 61.64, 60.97, 20.95, 20.81 (2C), 20.71, 20.66 (3C).

Thiophenyl 2,2′,3,3′,4′,6,6′-*O*-heptaacetyl-β-D-cellobioside:[293] Acetobromocellobiose (14.69 g, 21 mmol) was dissolved in CH_2Cl_2 (150 mL) and thiophenol (6.4 mL, 63 mmol) was added. A sodium bicarbonate buffer (1 M, pH = 9.4; 150 mL) supplemented with tetrabutylammonium hydrogensulfate (7.14 g, 21 mmol) was added to the organic mixture and the solution stirred vigorously in open atmosphere. Reaction was complete after 45 min, and the product was extracted into CH_2Cl_2 (450 mL). The organic layer was washed with 10% K_2CO_3 (200 mL), water (2 X 150 mL), and brine (200 mL) then dried over anhydrous Na_2SO_4. The solvents were removed under reduced pressure, and the resulting off-white powder was recrystallized from hot EtOH to yield the title compound as white crystals (14 g, 90% yield).



Chemical Formula: $C_{32}H_{40}O_{17}S$
Exact Mass: 728.20
Molecular Weight: 728.71

^{1}H NMR (400 MHz, cdcl$_3$) δ 7.50 – 7.45 (m, 2H, arom Hs), 7.34 – 7.27 (m, 3H, arom Hs), 5.19 (t, *J* = 9.2 Hz, 1H), 5.14 (t, *J* = 9.3 Hz, 1H), 5.06 (t, *J* = 9.6 Hz, 1H) [H^{I}-2, H^{II}-2,4], 4.95 – 4.87 (m, 2H, H^{I}-3, H^{II}-3), 4.67 (d, *J* = 10.1 Hz, 1H, H^{I}-1), 4.56 (dd, *J* = 11.9, 1.6 Hz, 1H, H^{I}-6), 4.50 (d, *J* = 7.9 Hz, 1H, H^{II}-1), 4.37 (dd, *J* = 12.5, 4.3 Hz, 1H, H^{I}-6'), 4.10 (dd, *J* = 11.9, 5.5 Hz, 1H, H^{II}-6), 4.03 (dd, *J* = 12.4, 2.1 Hz, 1H, H^{II}-6'), 3.73 (t, *J* = 9.5 Hz, 1H, H^{I}-4), 3.68 – 3.60 (m, 2H, H^{I}-5, H^{II}-5), 2.11 (s, 3H), 2.08 (s, 3H), 2.07 (s, 3H), 2.02 (d, *J* = 3.1 Hz, 3H), 2.01 (s, 6H), 1.98 (s, 3H) [Ac Hs]. ^{13}C NMR (101 MHz, cdcl$_3$) δ 170.41, 170.14 (2C), 169.66, 169.46, 169.22, 168.95 (Ac C=Os), 133.02 (2C), 131.69, 128.83 (2C), 128.26 (arom Cs), 100.70 (C^{II}-1), 85.47 (C^{I}-1), 76.29, 73.55, 72.87, 71.93, 71.53, 70.13, 67.73, 61.95, 61.48 (pyranose Cs), 20.78, 20.71, 20.58, 20.48 (4C) (Ac-Cs).

5.6 Methonium binding thermodynamics

5.6.1 Ligands 4.1a-e and 4.2a-d syntheses

Ligands 4.1a-e and 4.2a-d were synthesized and purified by Yi Wang as described by Wang et al[213] and detailed in Yi Wang's dissertation.[214]

5.6.2 Determination of Δδ from 1H-NMR binding studies of CB[7]•4.1a-d and CB[7]•4.2a-d ligands

The ^{1}H-NMR spectra of D_2O solutions containing roughly 8 mM CB[7] and 16 mM ligand was recorded on a Bruker 500 MHz NMR spectrometer. The exact ($[CB7]_{total}$ / $[Lig]_{total}$) ratio was determined by peak integration. Complexation-induced shifts (Δδ) were calculated based on the chemical shifts of the free ligand and equilibrium binding constant (K_a) measured with ITC. The observed chemical shift of a proton in the fast exchange regime (δ_{obs}) is a linear combination of the chemical shifts of the bound form (δ_b) and the free form (δ_f) based on their relative population (f):

$$\delta_{obs} = f_b\delta_b + f_f\delta_f$$

Equation 47

The relative populations between the bound and free form can be used to eliminate the unknown chemical shift of bound form (δ_b):

$$f_f = 1 - f_b$$

Equation 48

$$\delta_{obs} = f_b(\delta_b - \delta_f) + \delta_f$$

Equation 49

Thus, the shielding/deshielding effect (Δδ) can be obtained using only known information regarding the chemical shifts of the free ligand and the observed chemical shift in fast exchange:

$$\Delta\delta = \delta_b - \delta_f$$

Equation 50

$$\Delta\delta = \frac{\delta_{obs} - \delta_f}{f_b}$$

Equation 51

With the knowledge of binding constants (see Chapter 4), it is straightforward to calculate the equilibrium concentration of free ligand and thus determine f_b. Due to the high binding constants of ligands investigated in the current study, f_b is very similar to the ratio between total receptor concentration and total ligand concentration (CB[7]/ligand ratios are provided in Figure 44 through Figure 52), indicative of the saturation of receptor binding sites. The exact ratio ($[CB7]_{total}$ / $[Lig]_{total}$) is determined by integration of the peak areas in the complex spectra. Despite the inherent errors with this approach, we are confident that the resulting f_b determined with this method is valid, given the consistency among results from a variety of CB[7]-ligand complexes.

Appendix A

Distribution fitting and binding probabilities of force spectroscopy histograms

Inverse Gaussian distribution functions from the Matlab™ fitting toolkit were used to fit the rupture force and length histograms generated from His_6, MCC, and G3 un-binding experiments. The mean and standard deviation are presented along with the fit and histograms in Figure 64-Figure 67. The experimental binding probabilities are given in Table 10.

Table 10: Binding probabilities from force spectroscopy experiments

Experiment	Initial P_{bind}	Blocked P_{bind}	Washed P_{bind}
MCC (no Tween)	0.93	0.55	0.73
MCC (0.01% Tween)	0.99	0.34	0.69
MCC (0.05% Tween)	0.89	0.34	0.76
G3 (exp1)	0.73	0.32	ND*
G3 (exp2)	0.75	0.45	0.596
Blank Tip	0.16	ND	ND

* ND: not determined.

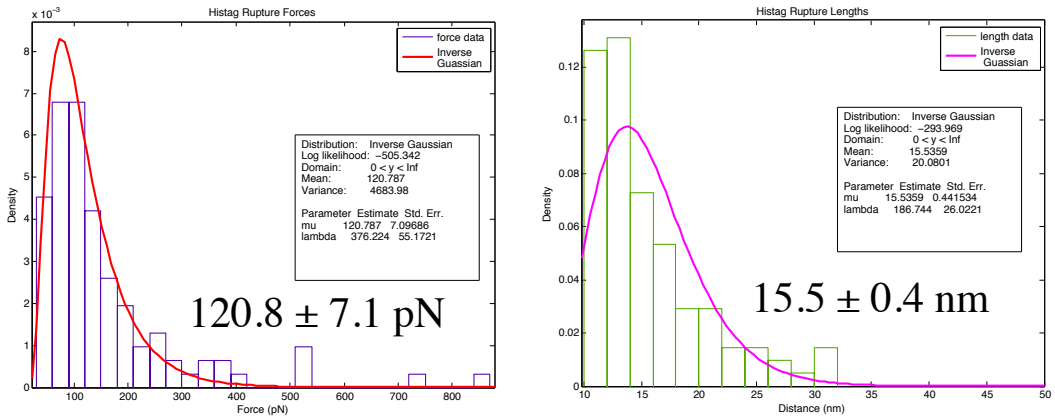


Figure 64: Inverse Gaussian fit to His6 rupture histograms

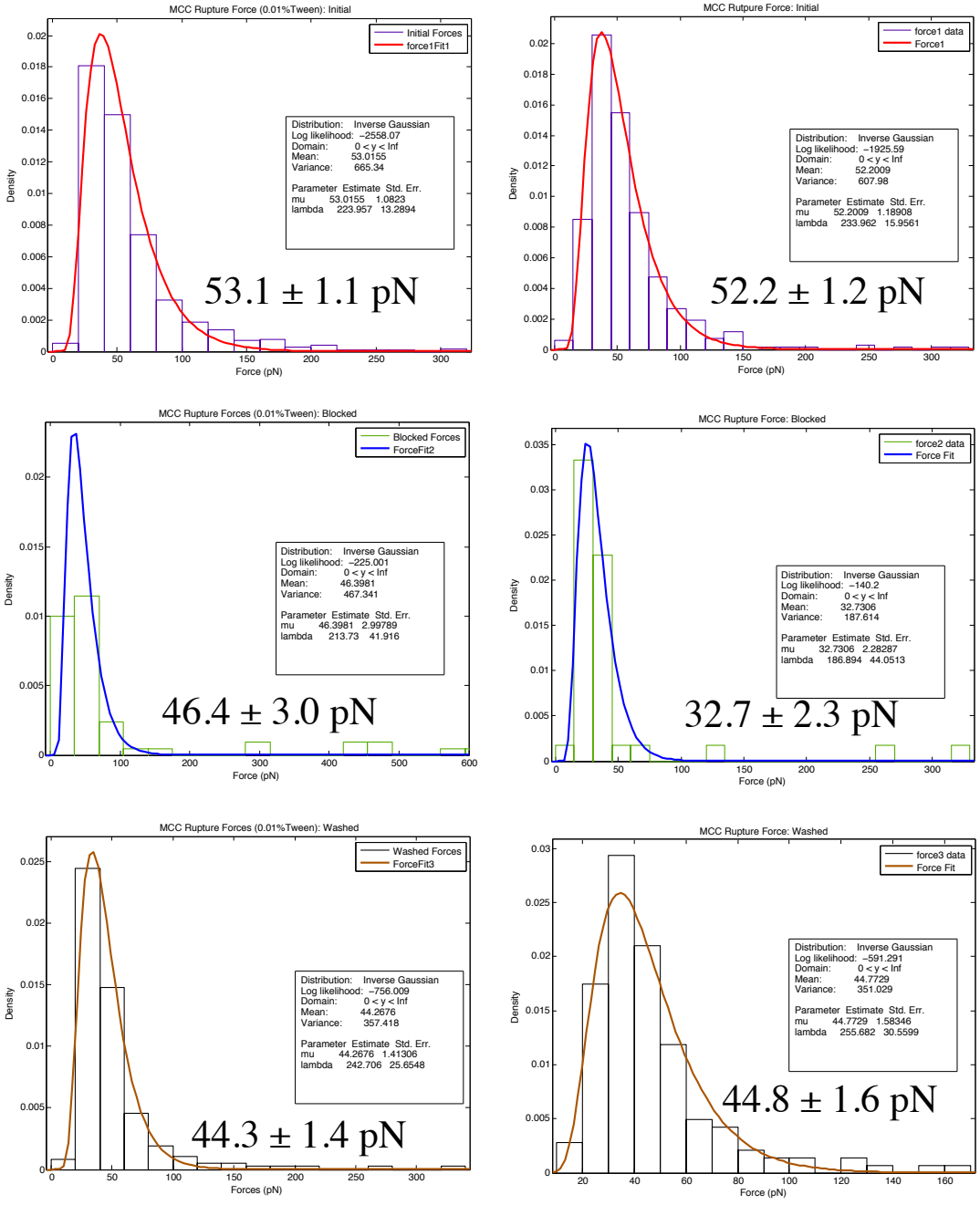


Figure 65: Inverse Gaussian fits to miniCipC-Cellulose rupture force histograms

*Data on the left is for 0.01% Tween-20, and on the right is 0.05% Tween.

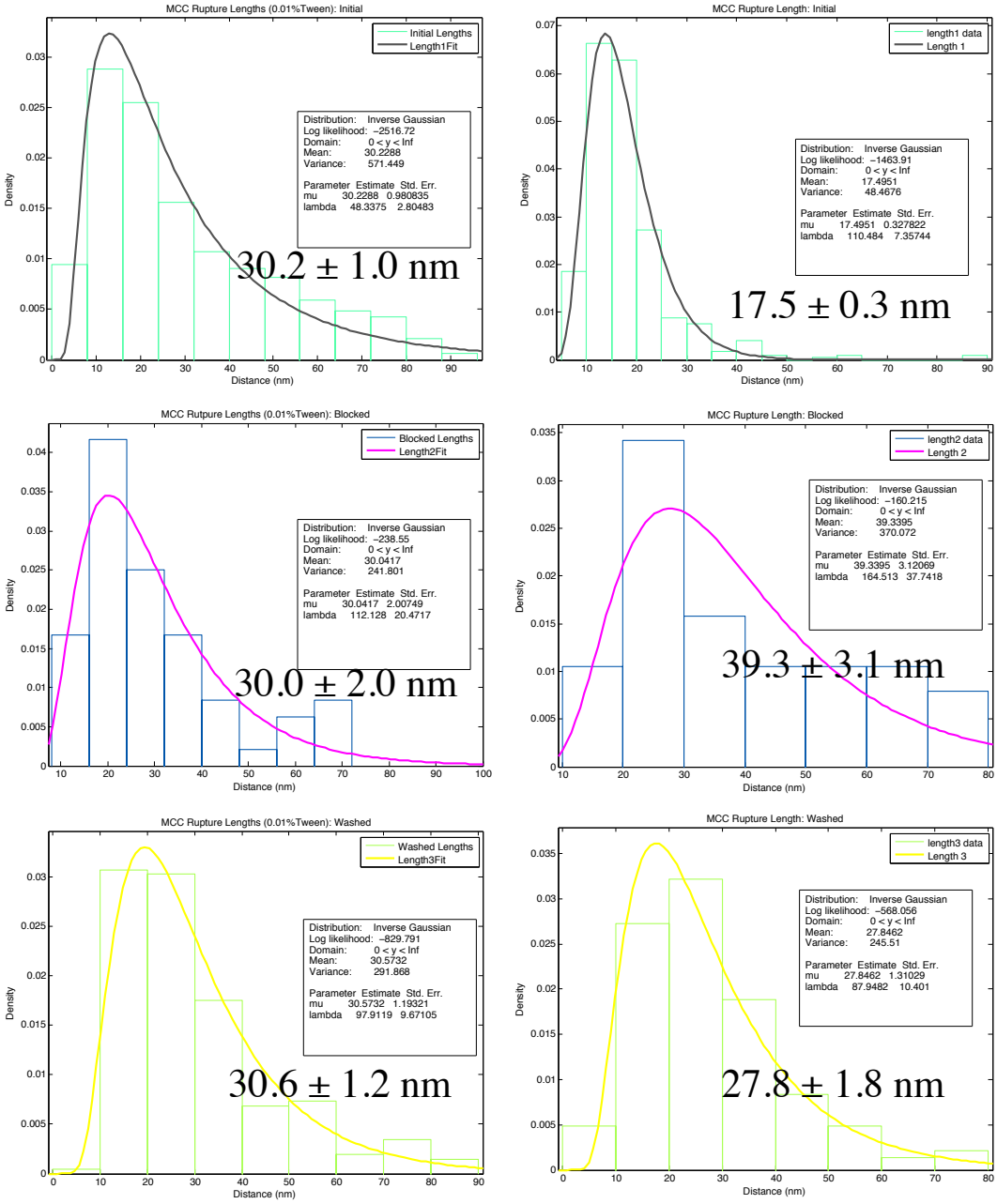


Figure 66: Inverse Guassian fits to miniCipC-cellulose rupture length histograms

*Data on the left is for 0.01% Tween-20, and on the right is 0.05% Tween.

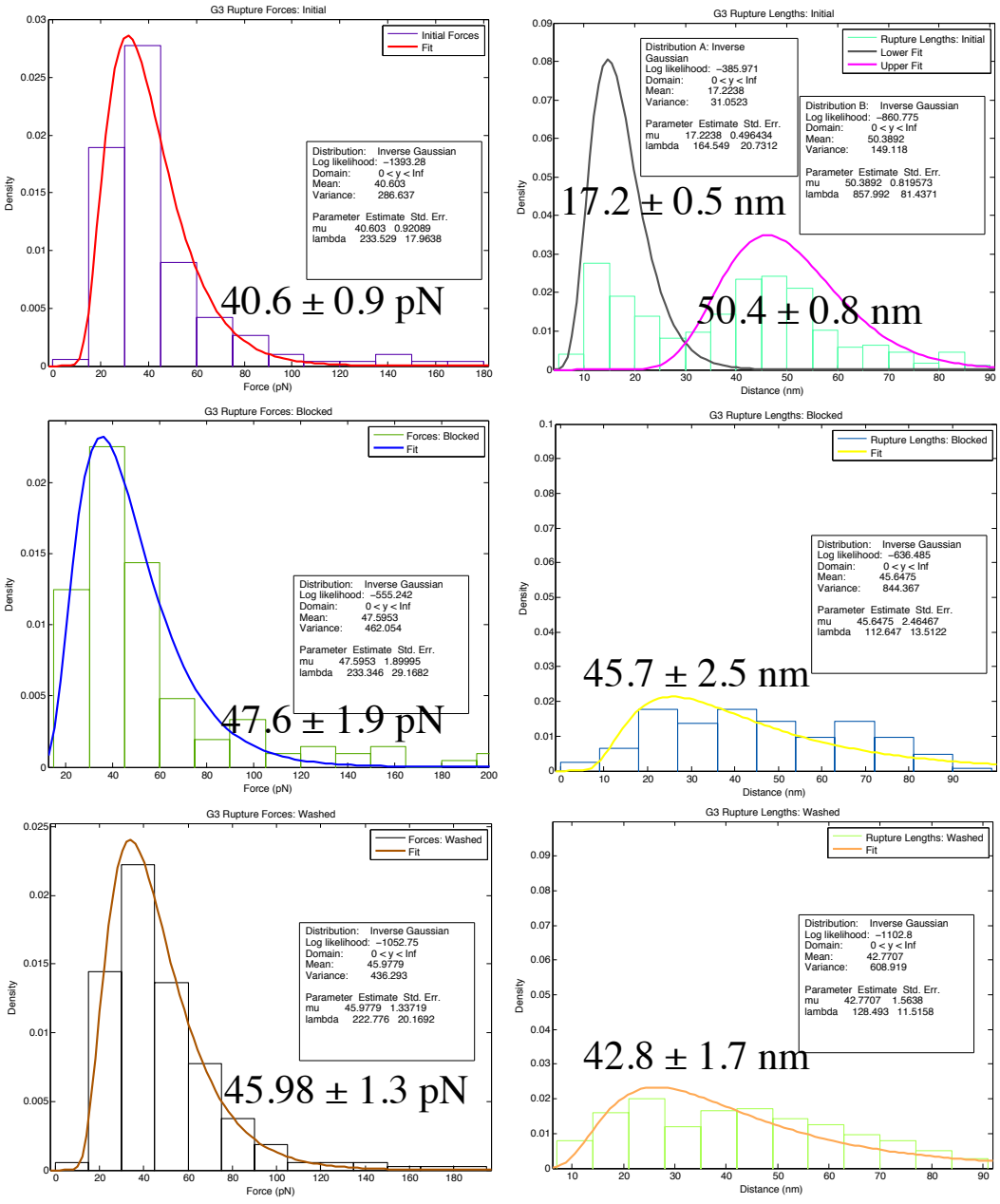


Figure 67: Inverse Guassian fits to Galectin 3-cellulose force and length histograms

*Data on the left is for forces, and on the right are lengths.

Appendix B

NMR spectra for novel compounds.

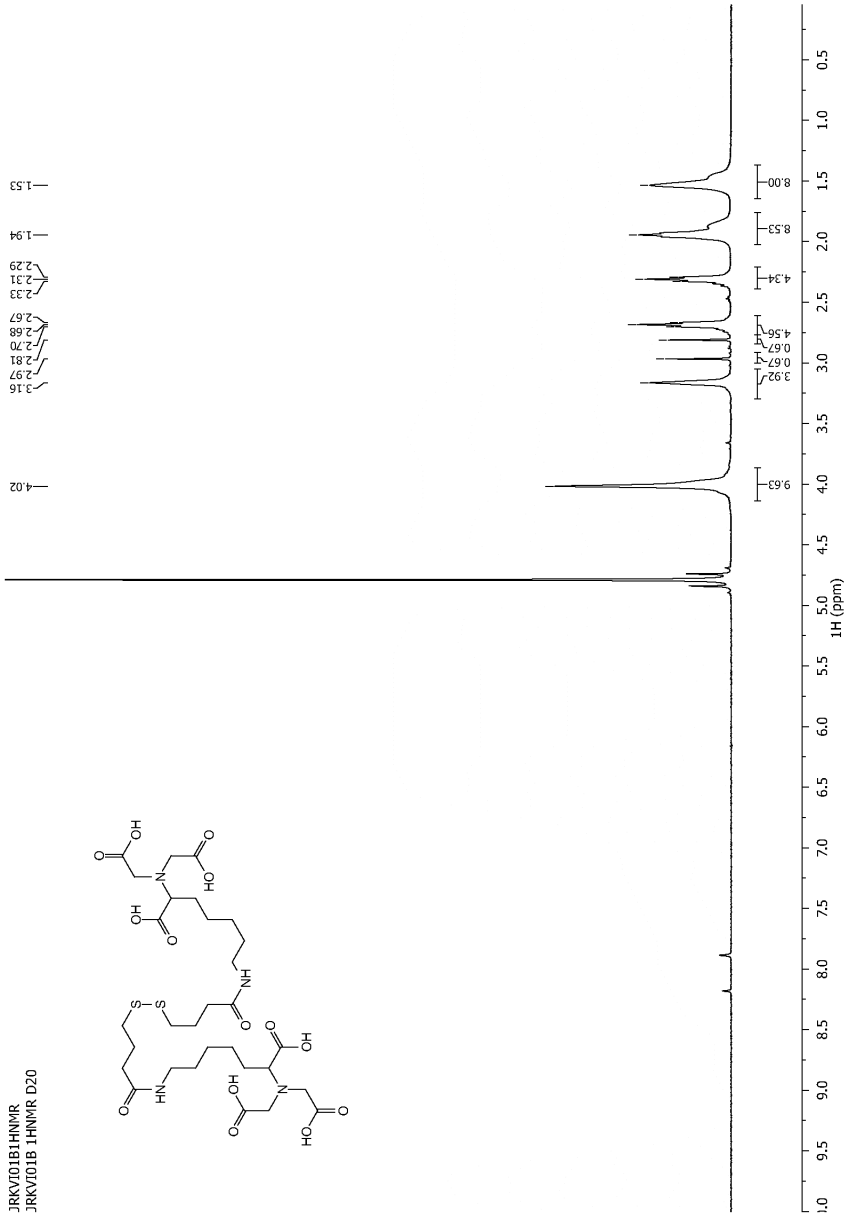


Figure 68: Compound 2.2 1H-NMR

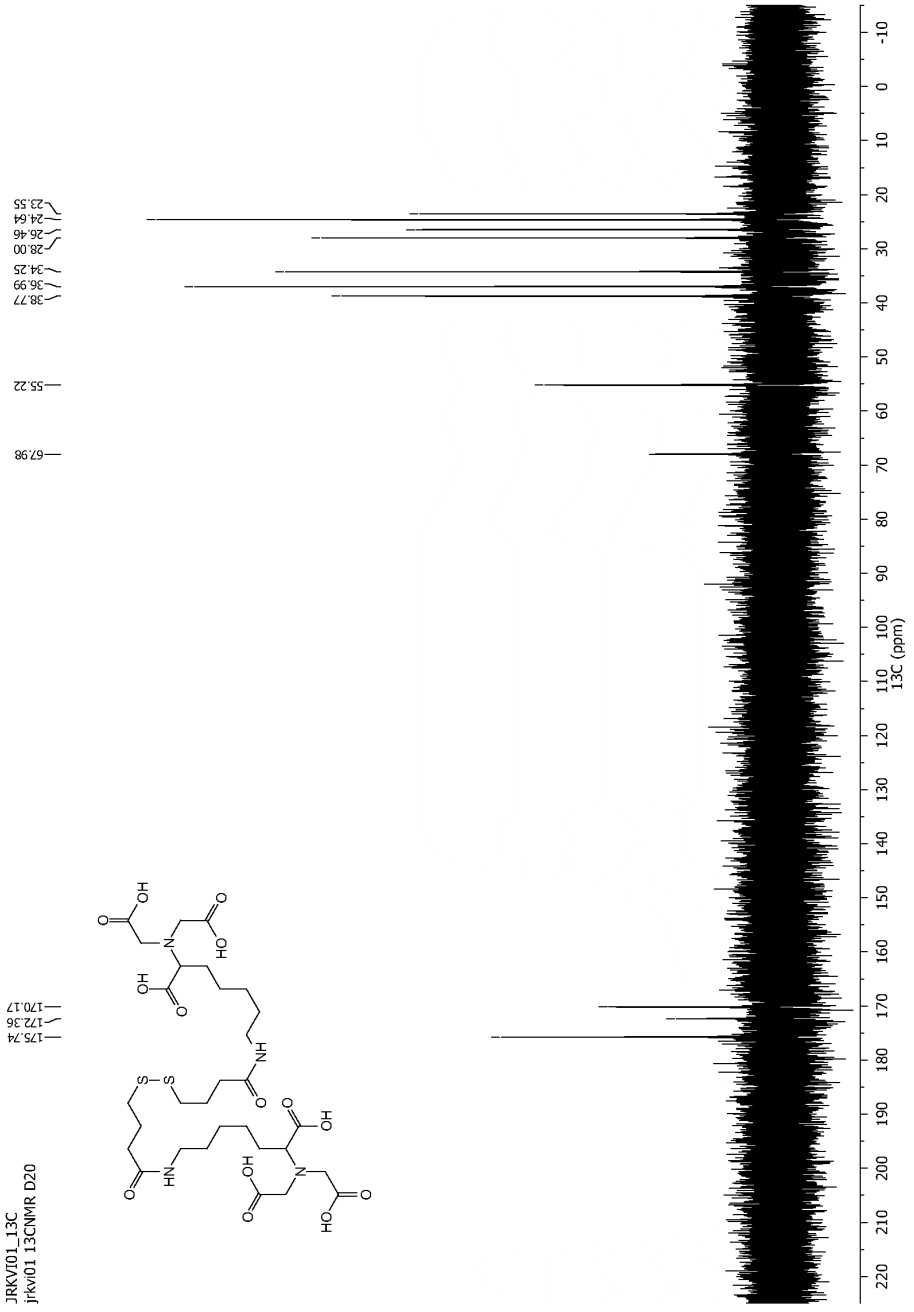


Figure 69: Compound 2.2 13C-NMR

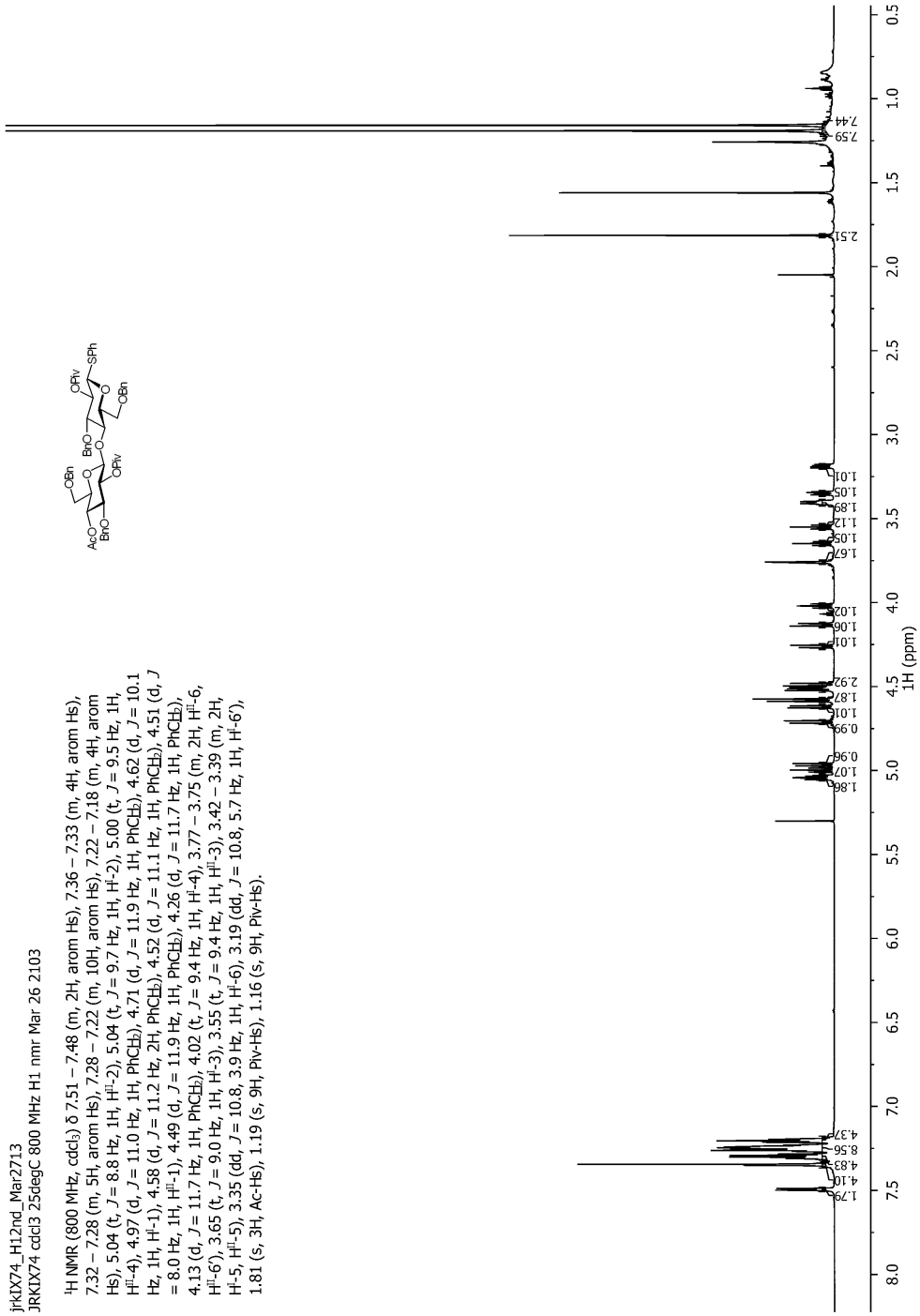


Figure 70: Compound 3.2 1H-NMR

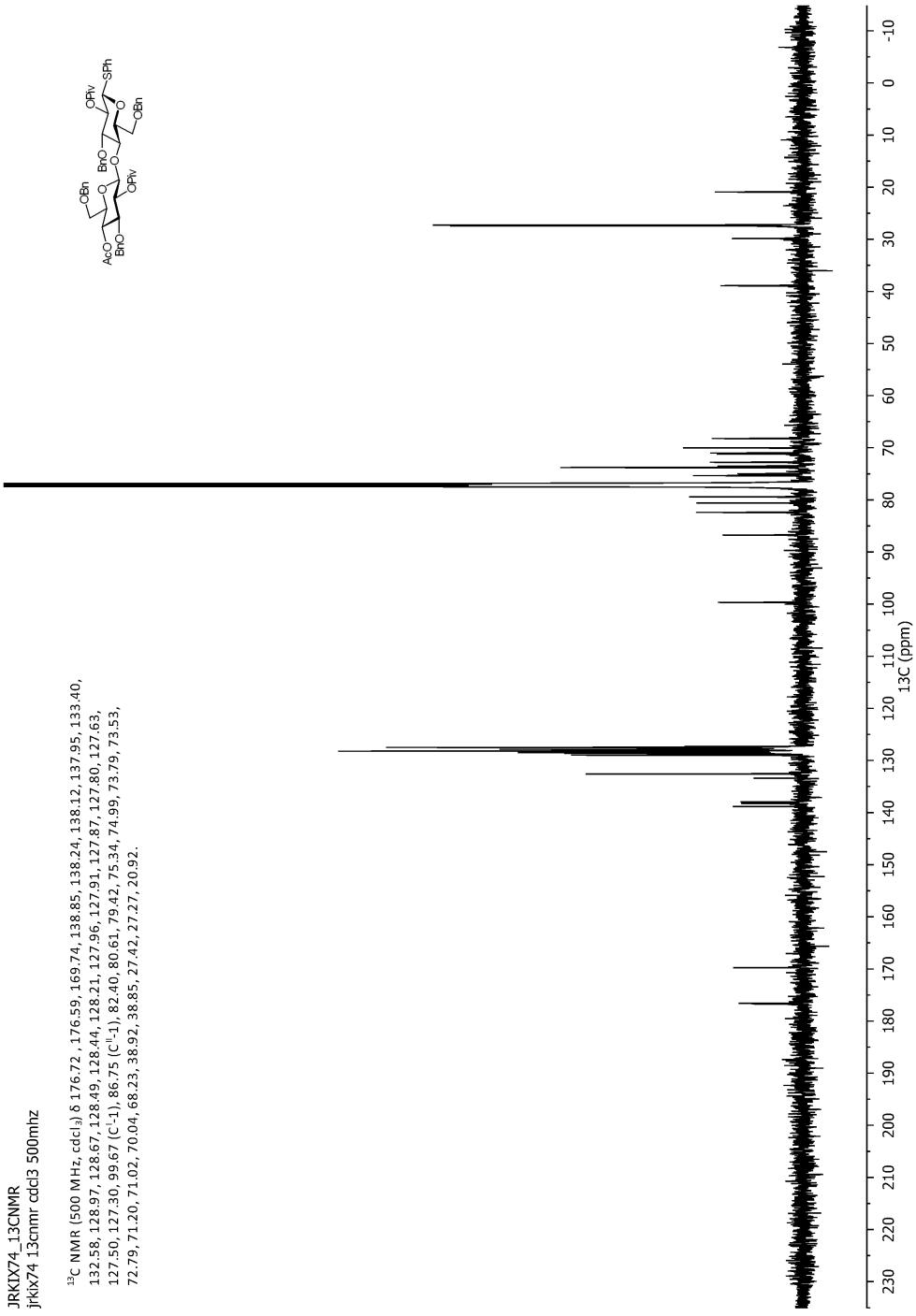


Figure 71: Compound 3.2 13C-NMR

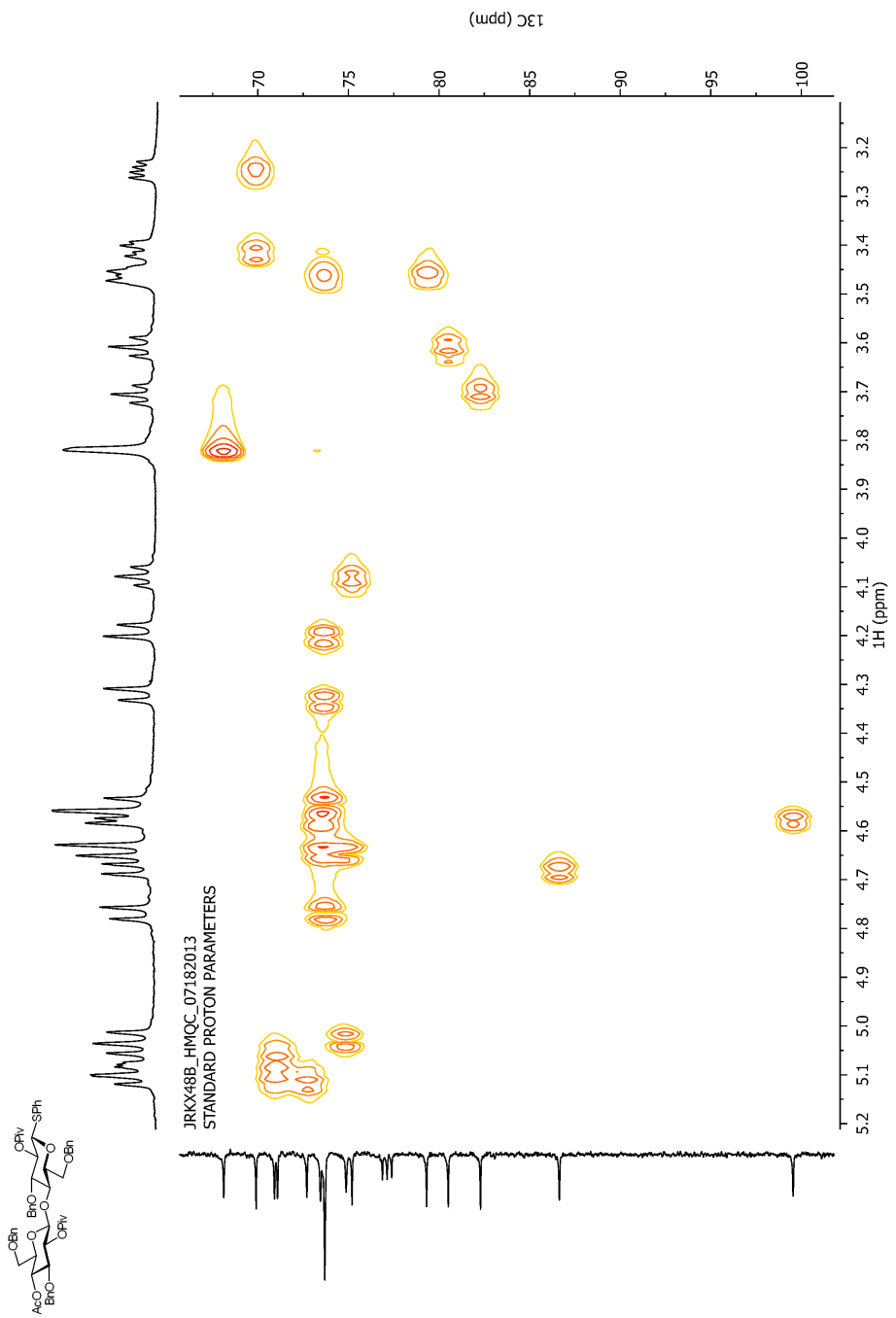


Figure 72: Compound 3.2 COSY

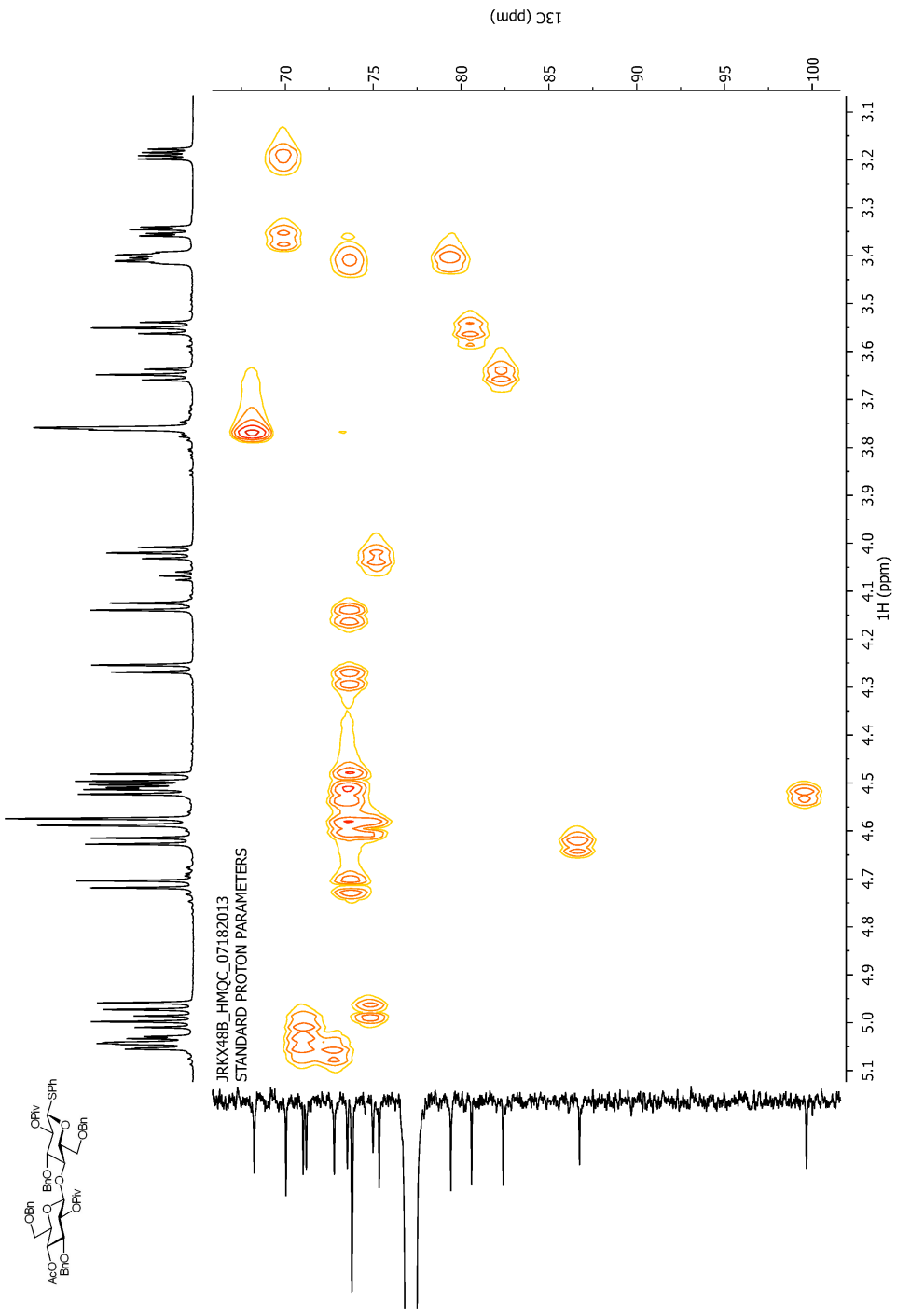


Figure 73: Compound 3.2 HMQC

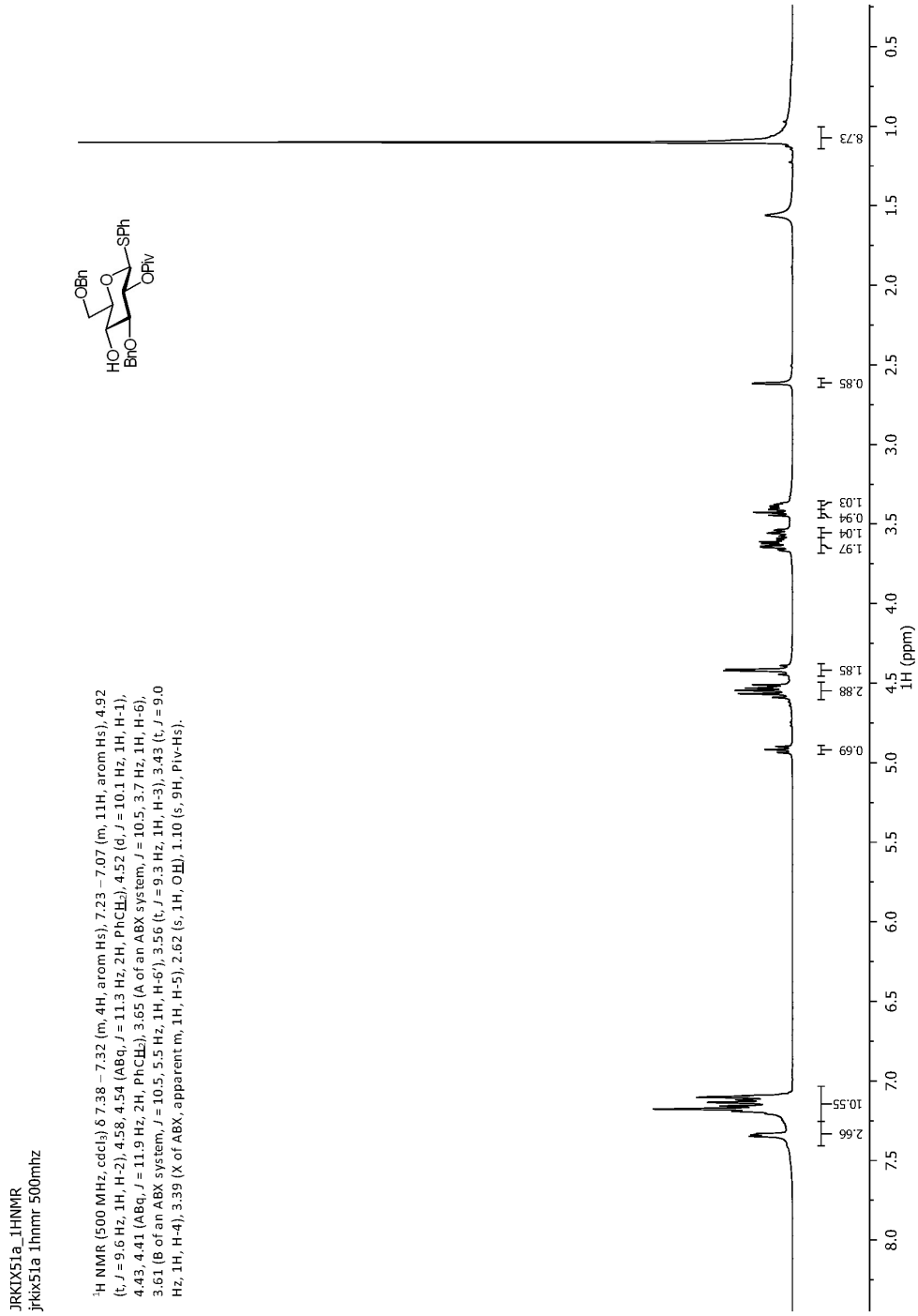


Figure 74: Compound 3.6 1H-NMR

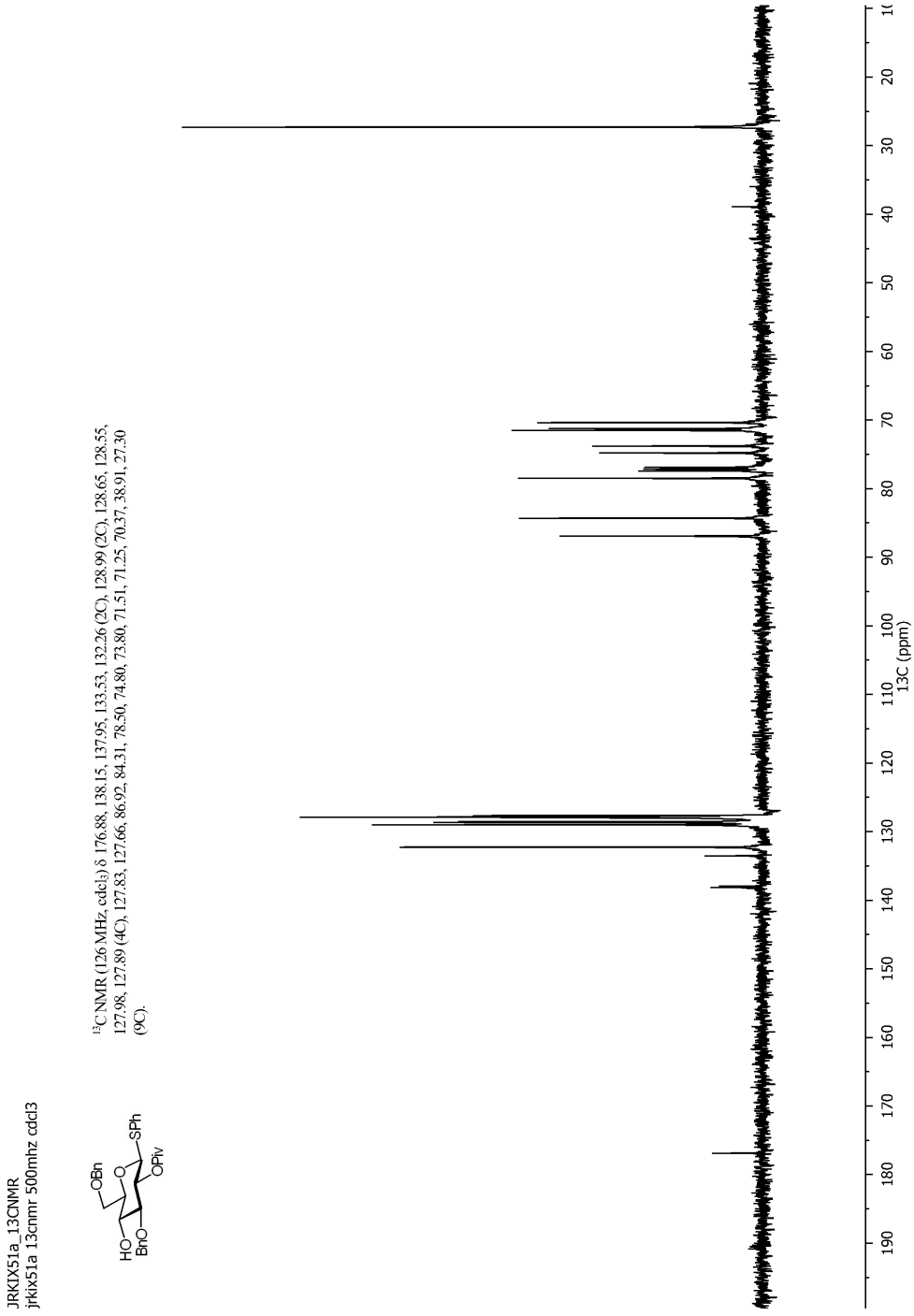


Figure 75: Compound 3.6 13C-NMR

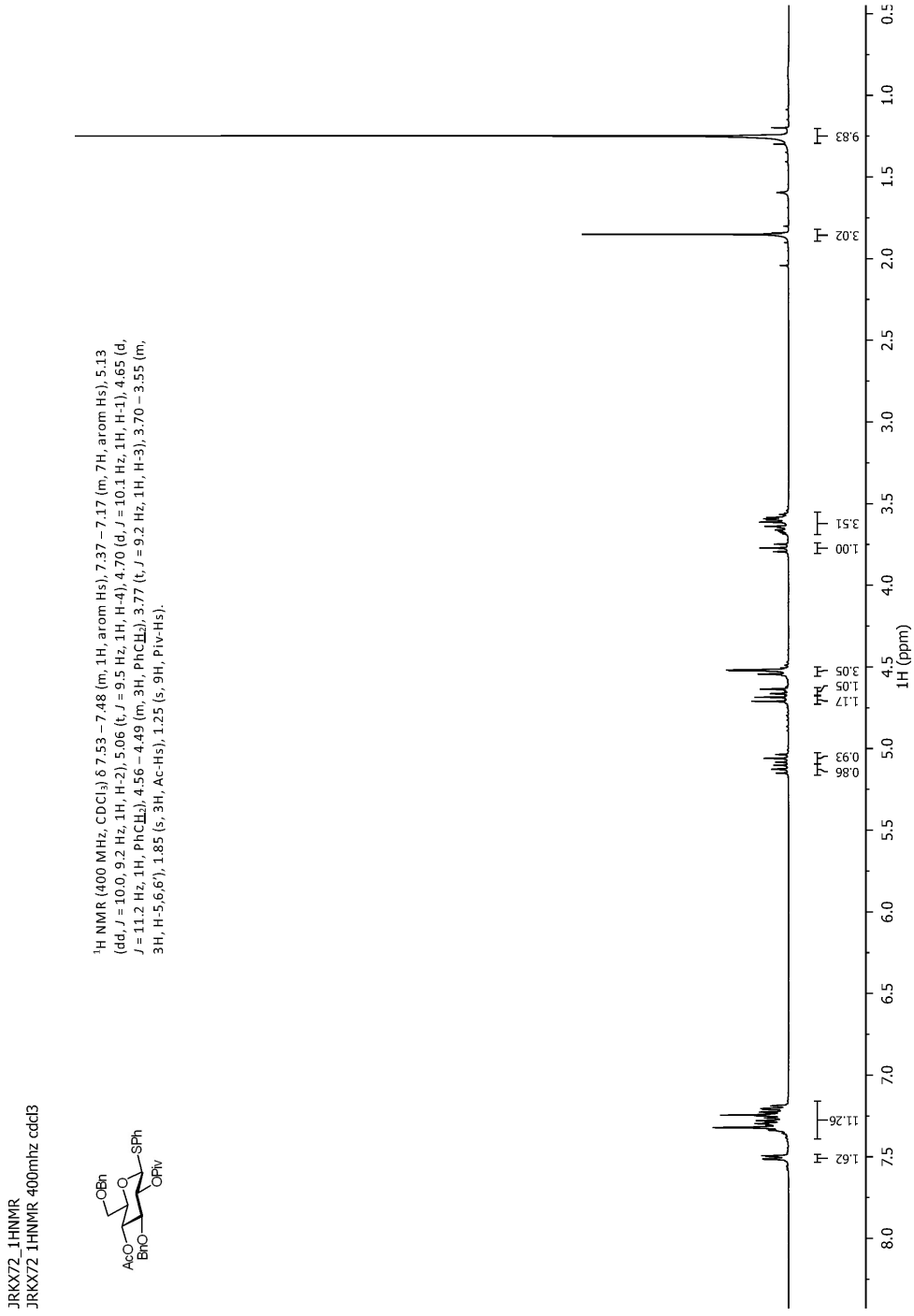


Figure 76: Compound 3.6a 1H-NMR

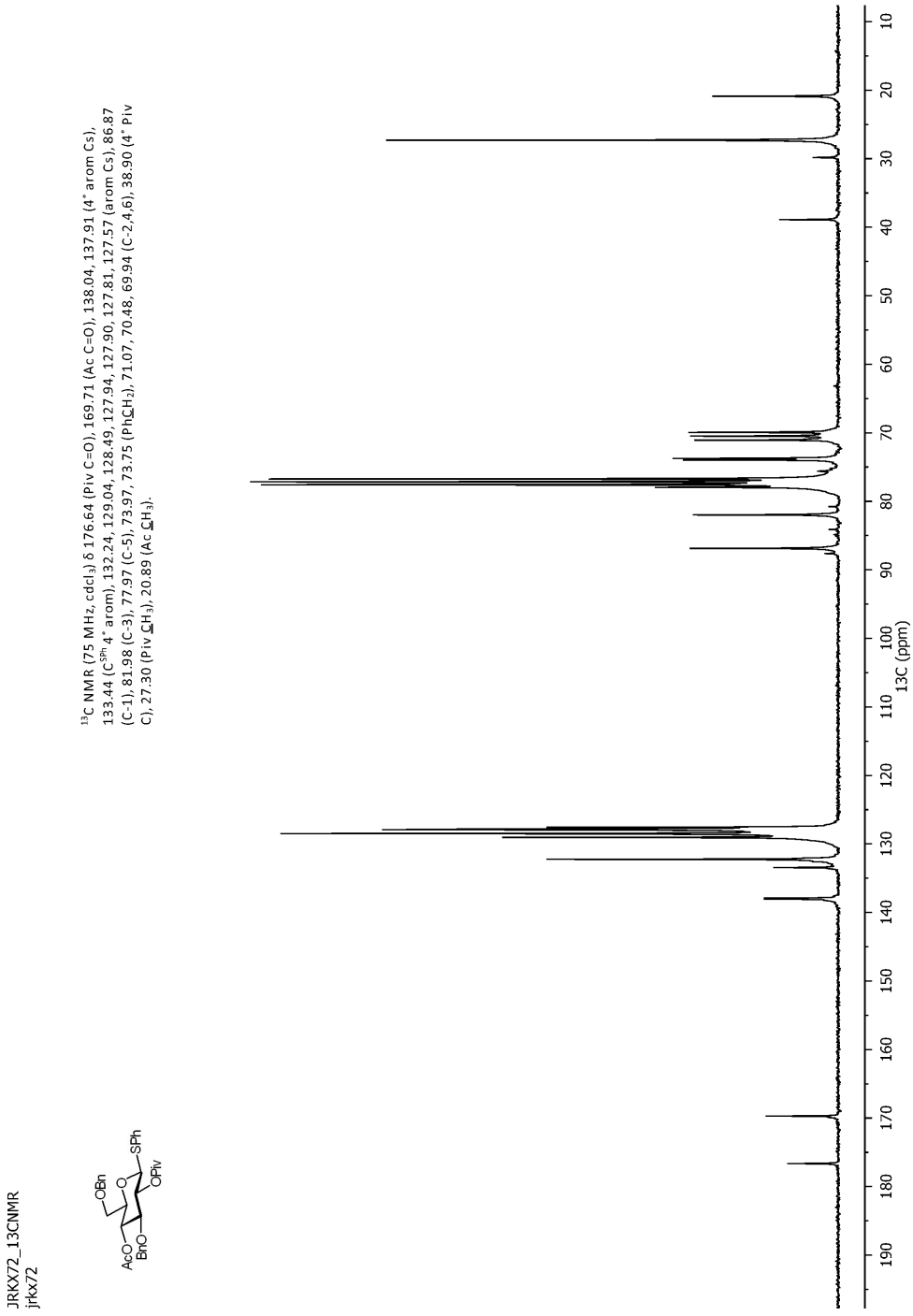


Figure 77: Compound 3.6a 13C-NMR

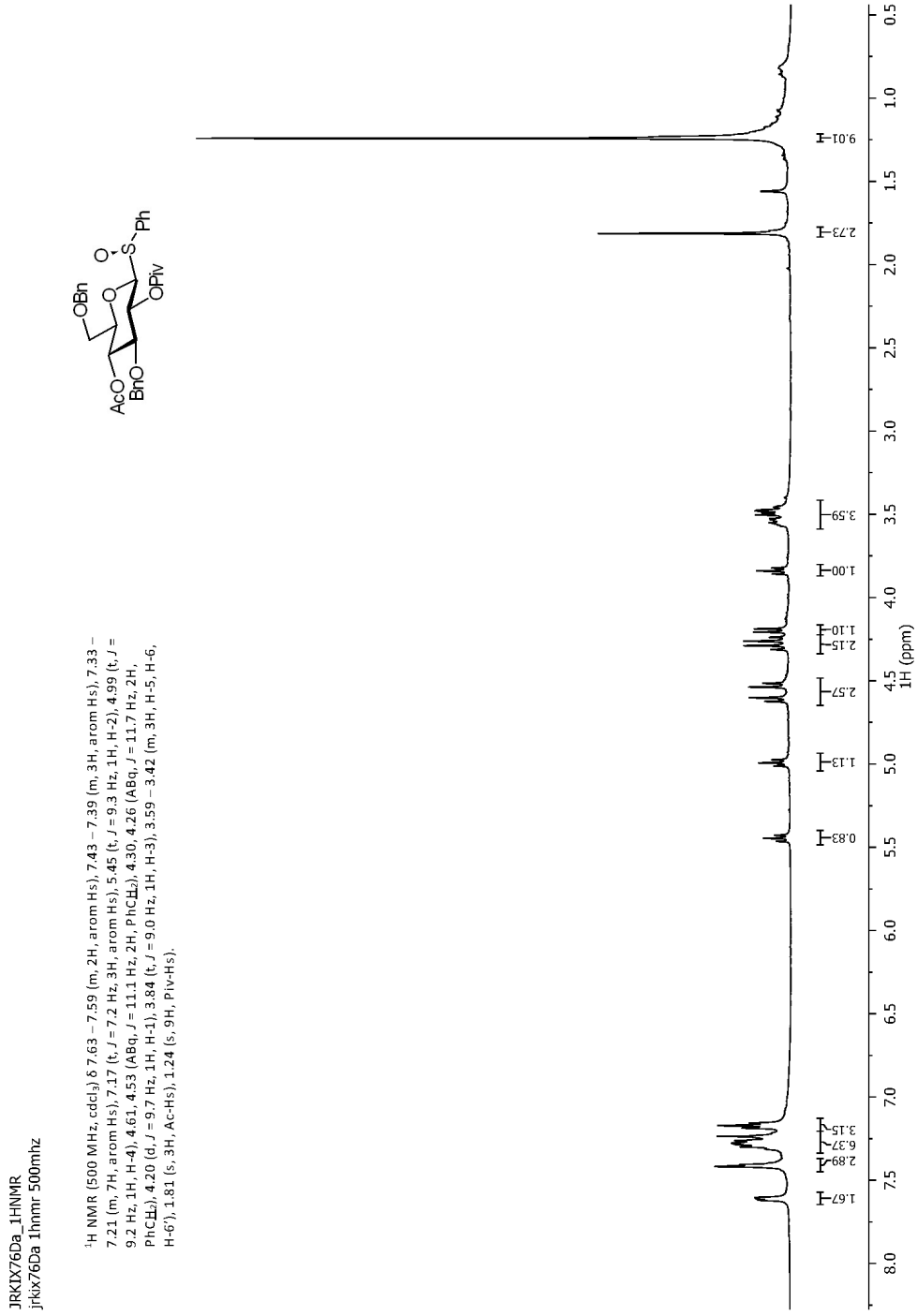


Figure 78: Compound 3.7a 1H-NMR

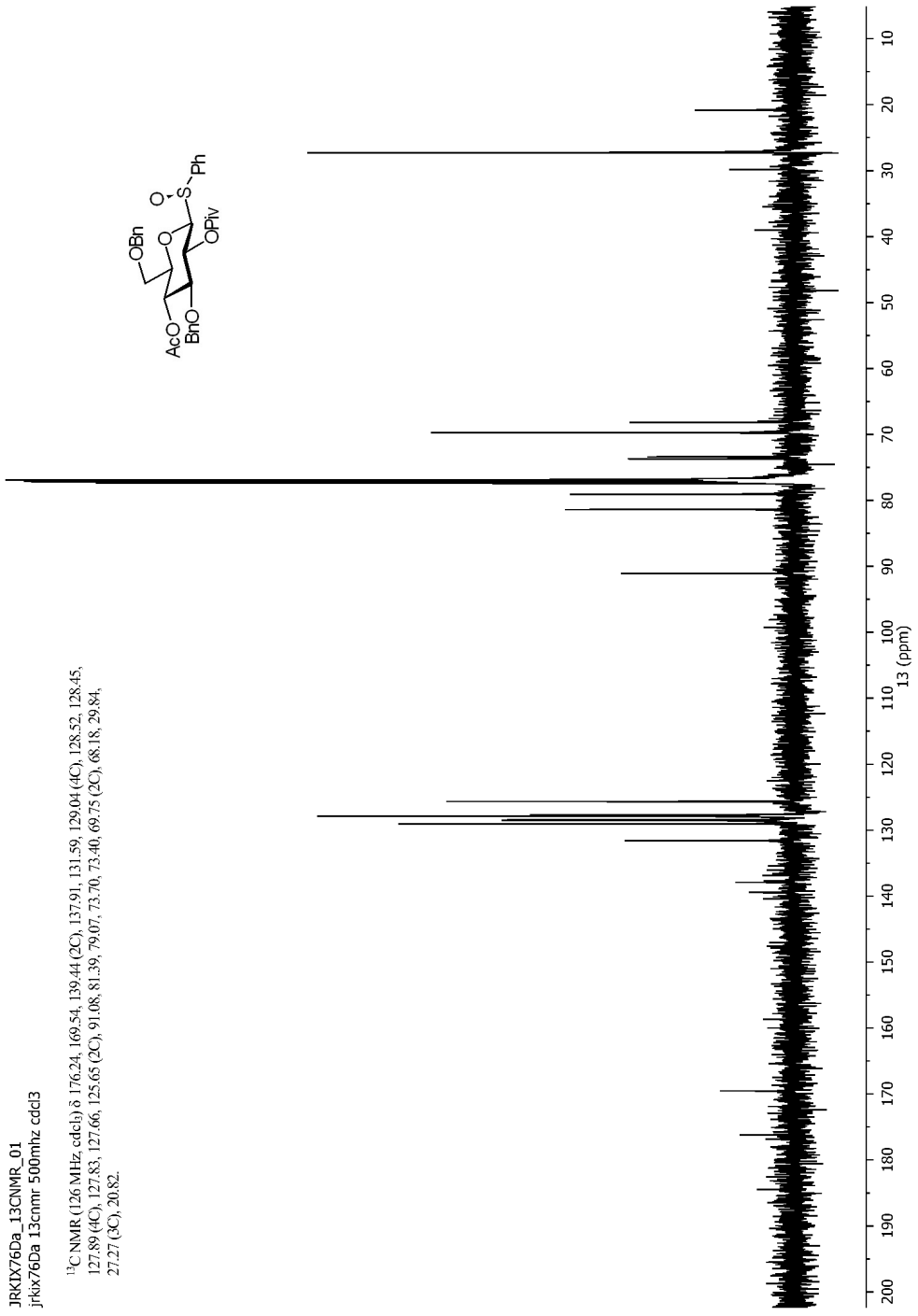


Figure 79: Compound 3.7a 13C-NMR

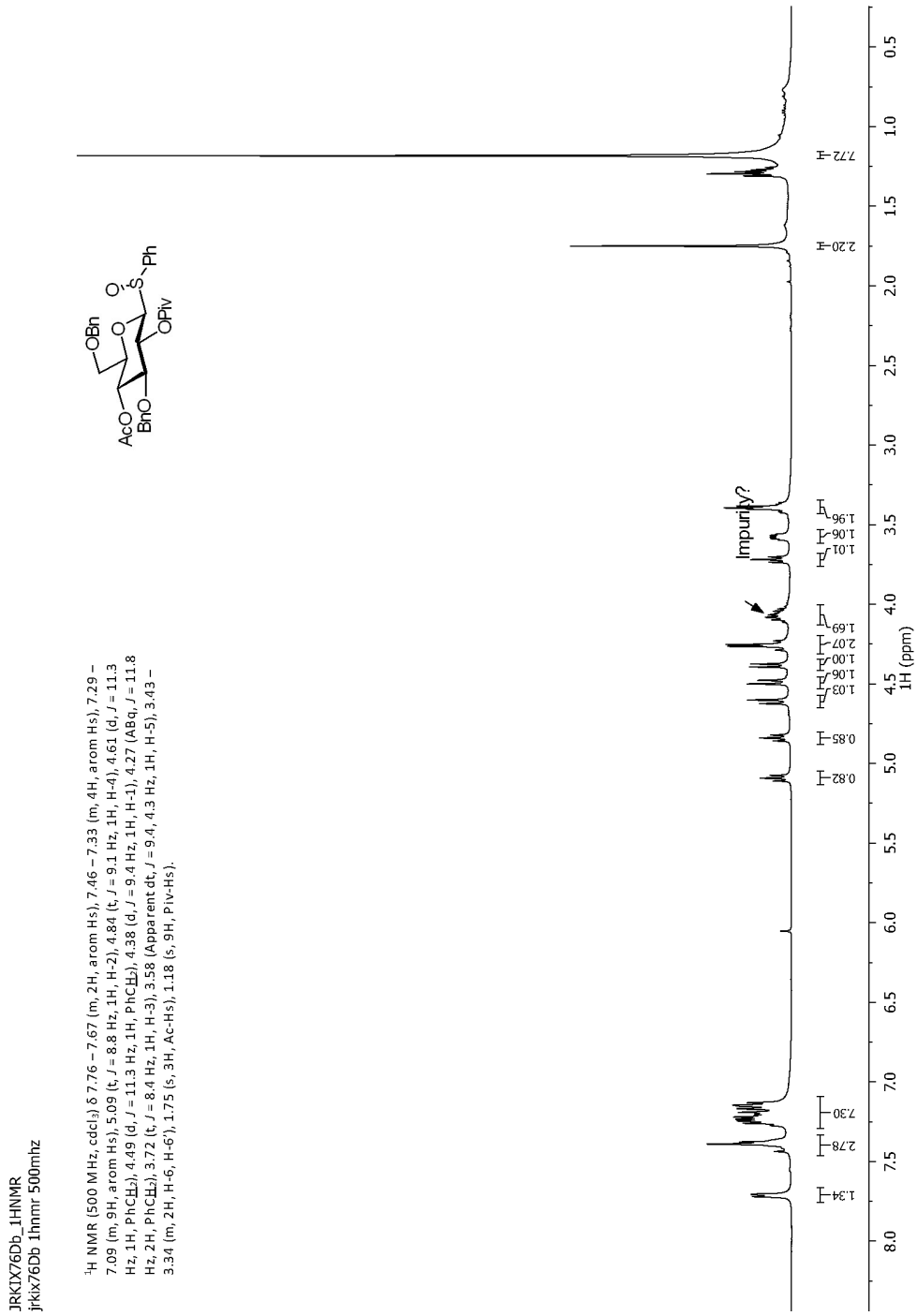


Figure 80: Compound 3.7b 1H-NMR

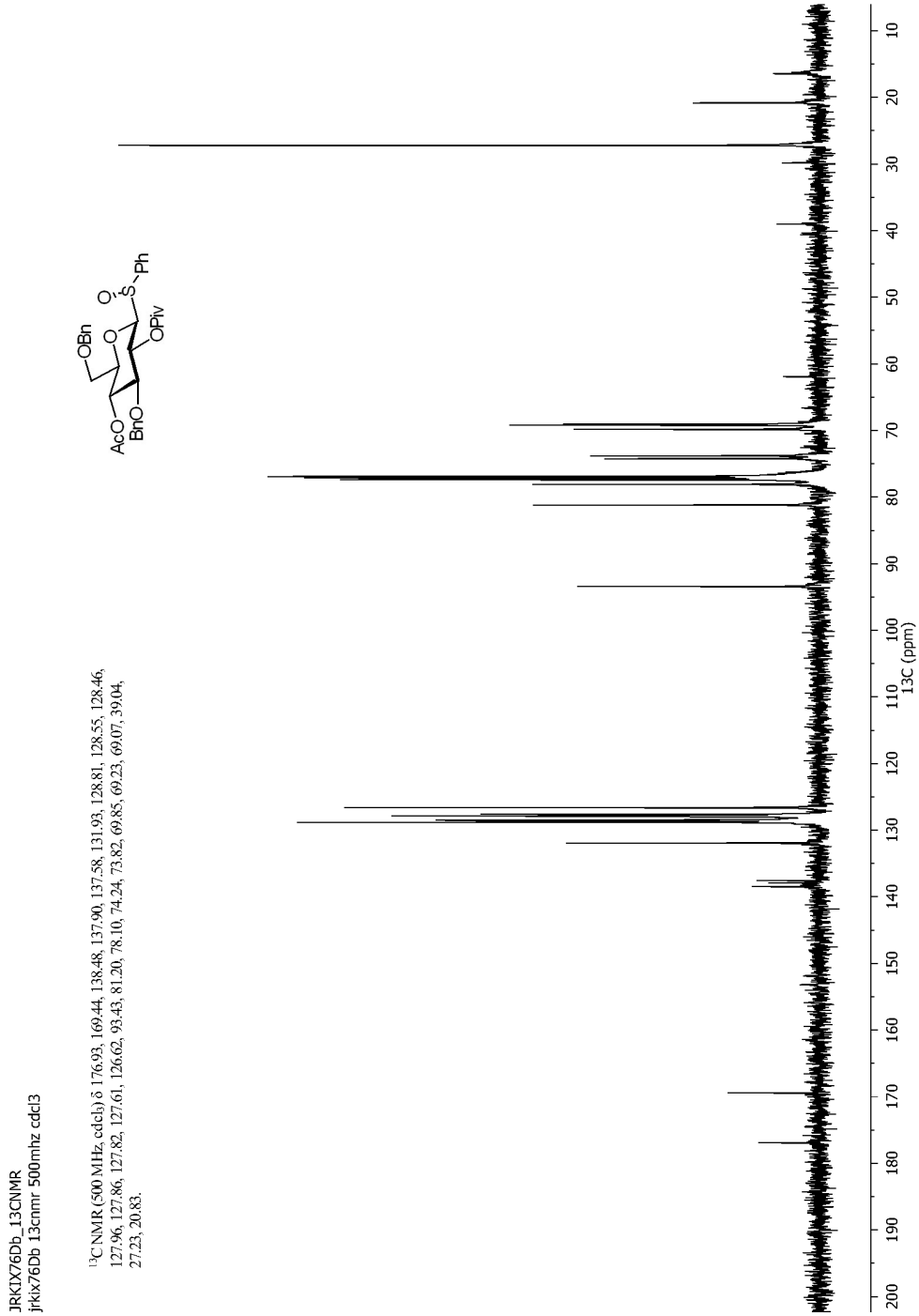


Figure 81: Compound 3.7b 13C-NMR

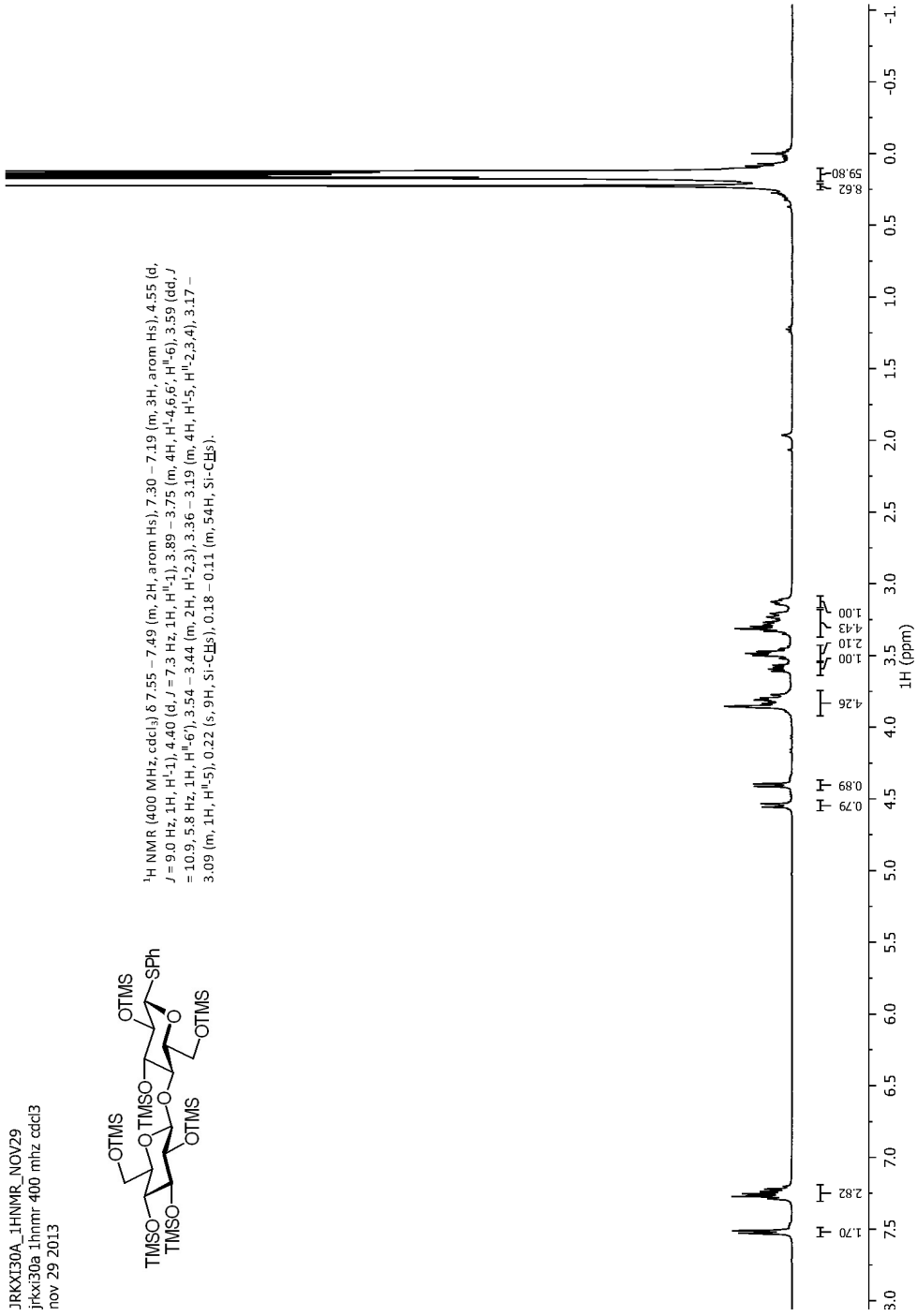


Figure 82: Compound 3.8 1H-NMR

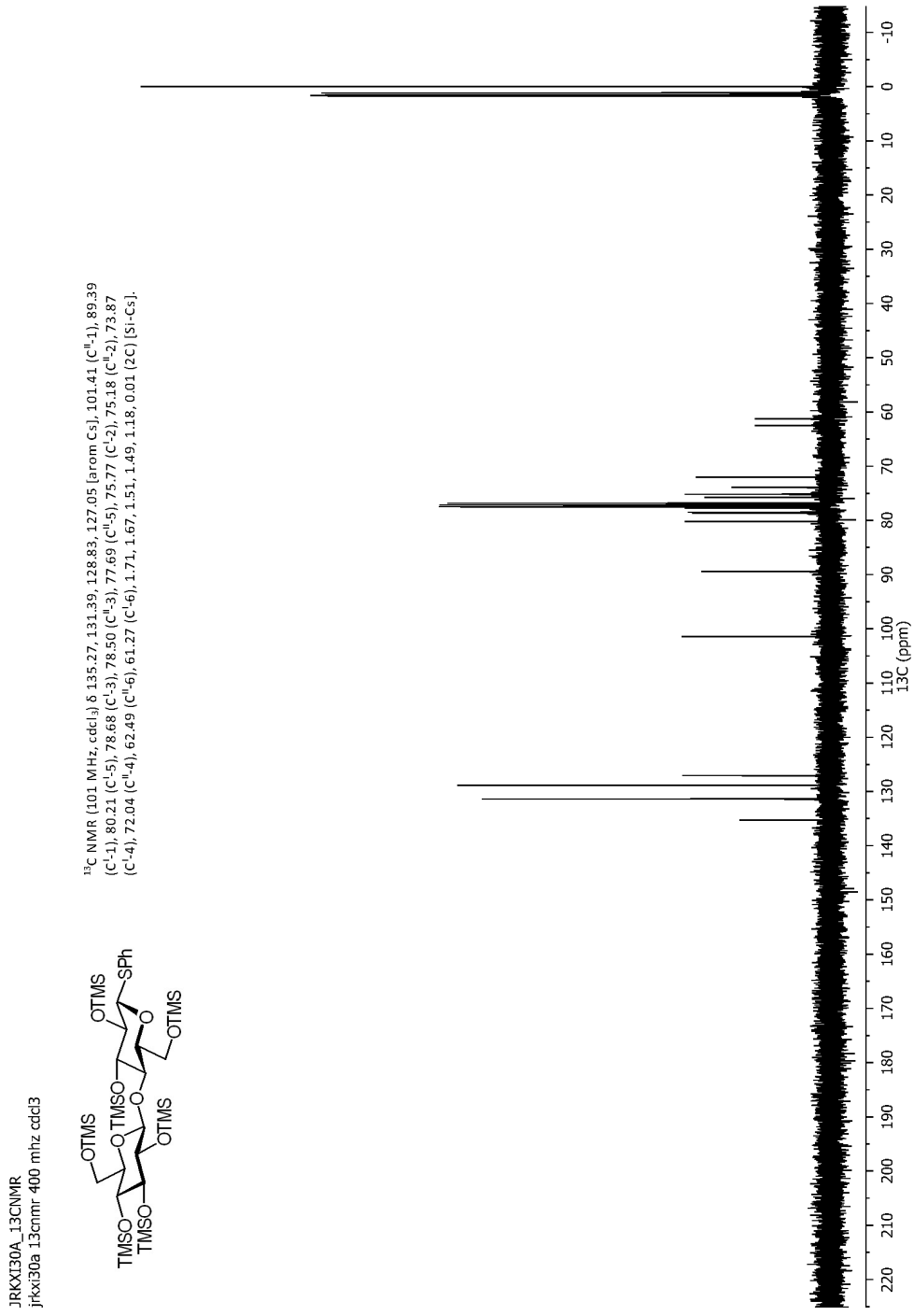


Figure 83: Compound 3.8 13C-NMR

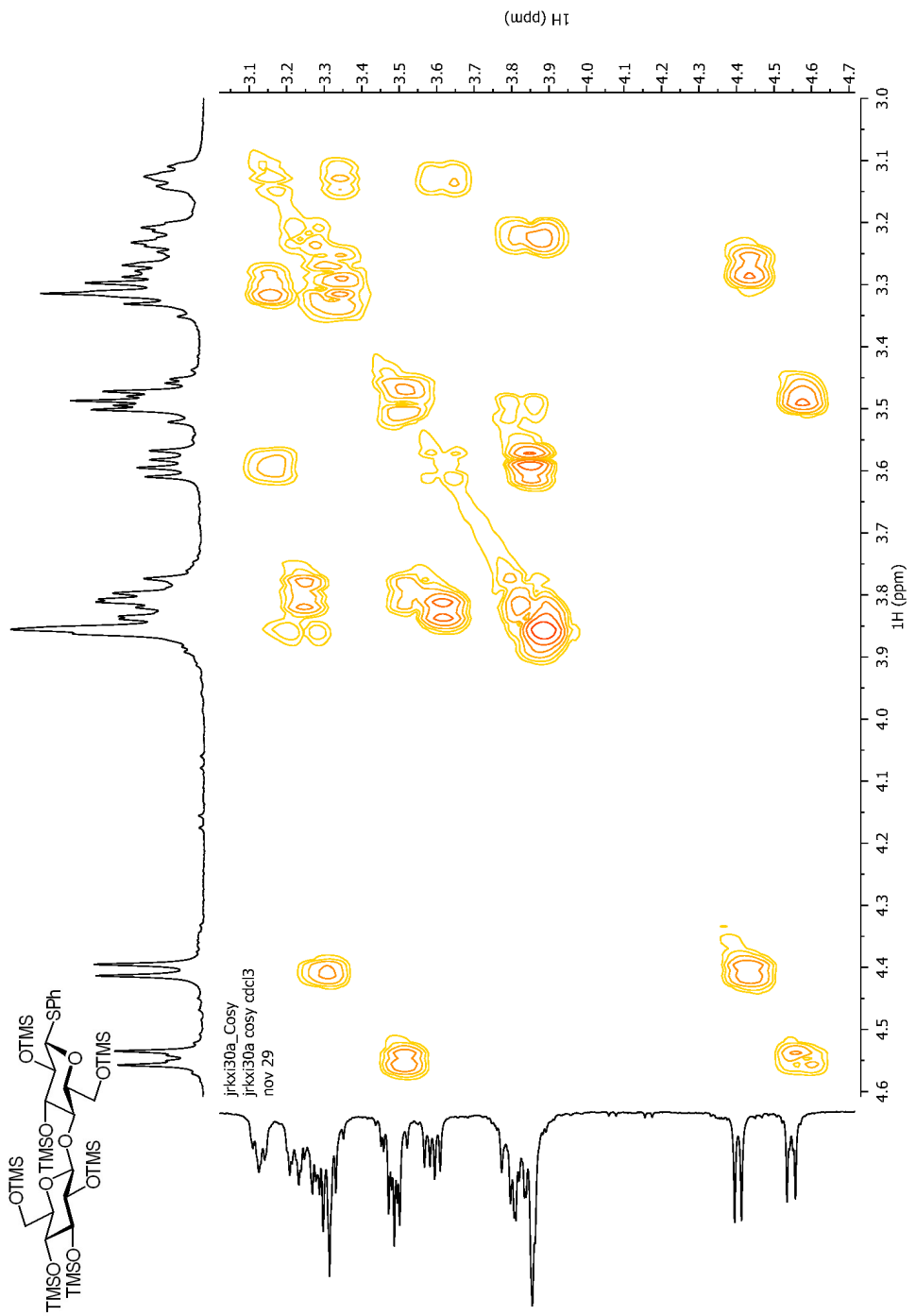


Figure 84: Compound 3.8 COSY

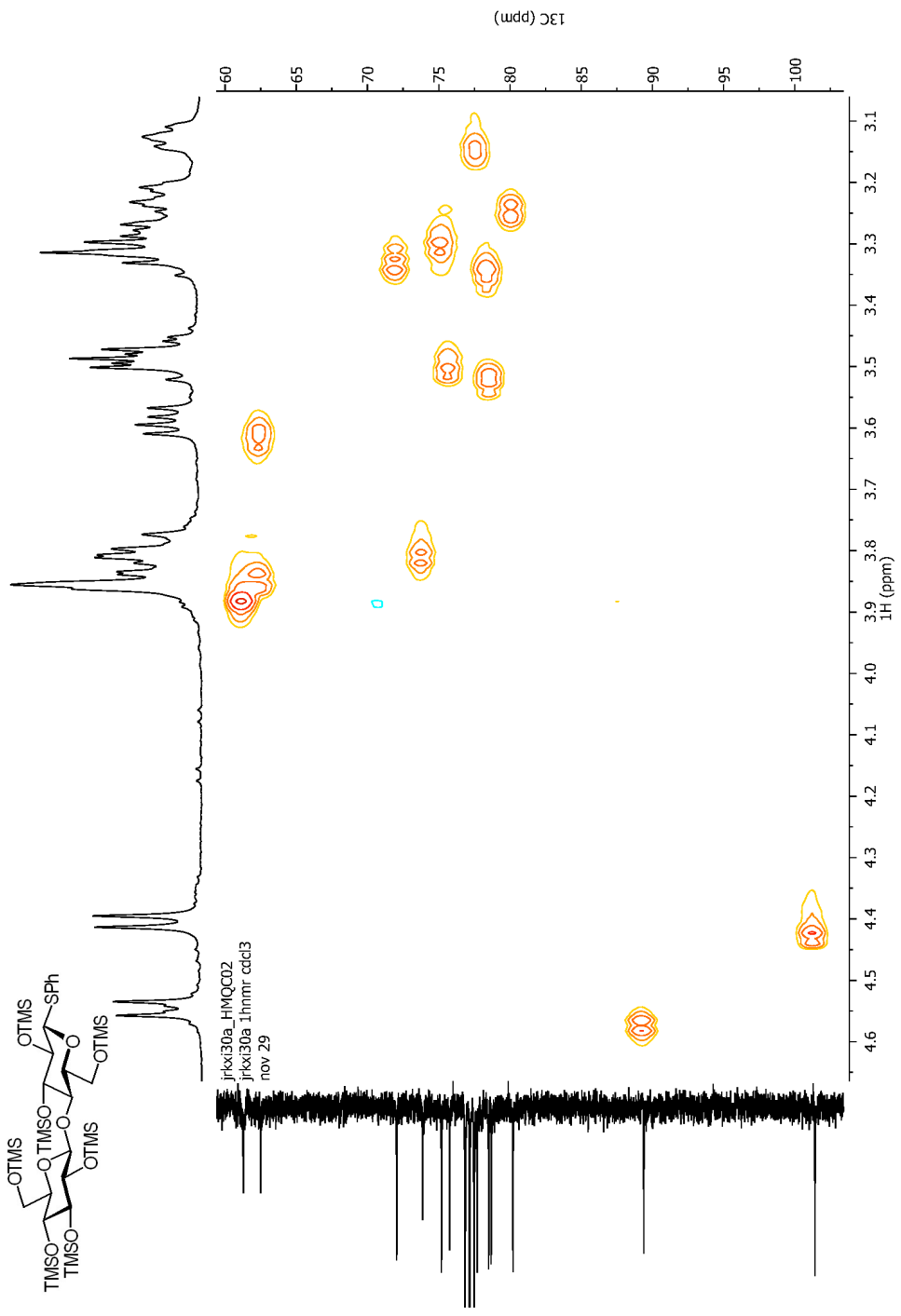


Figure 85: Compound 3.8 HMQC

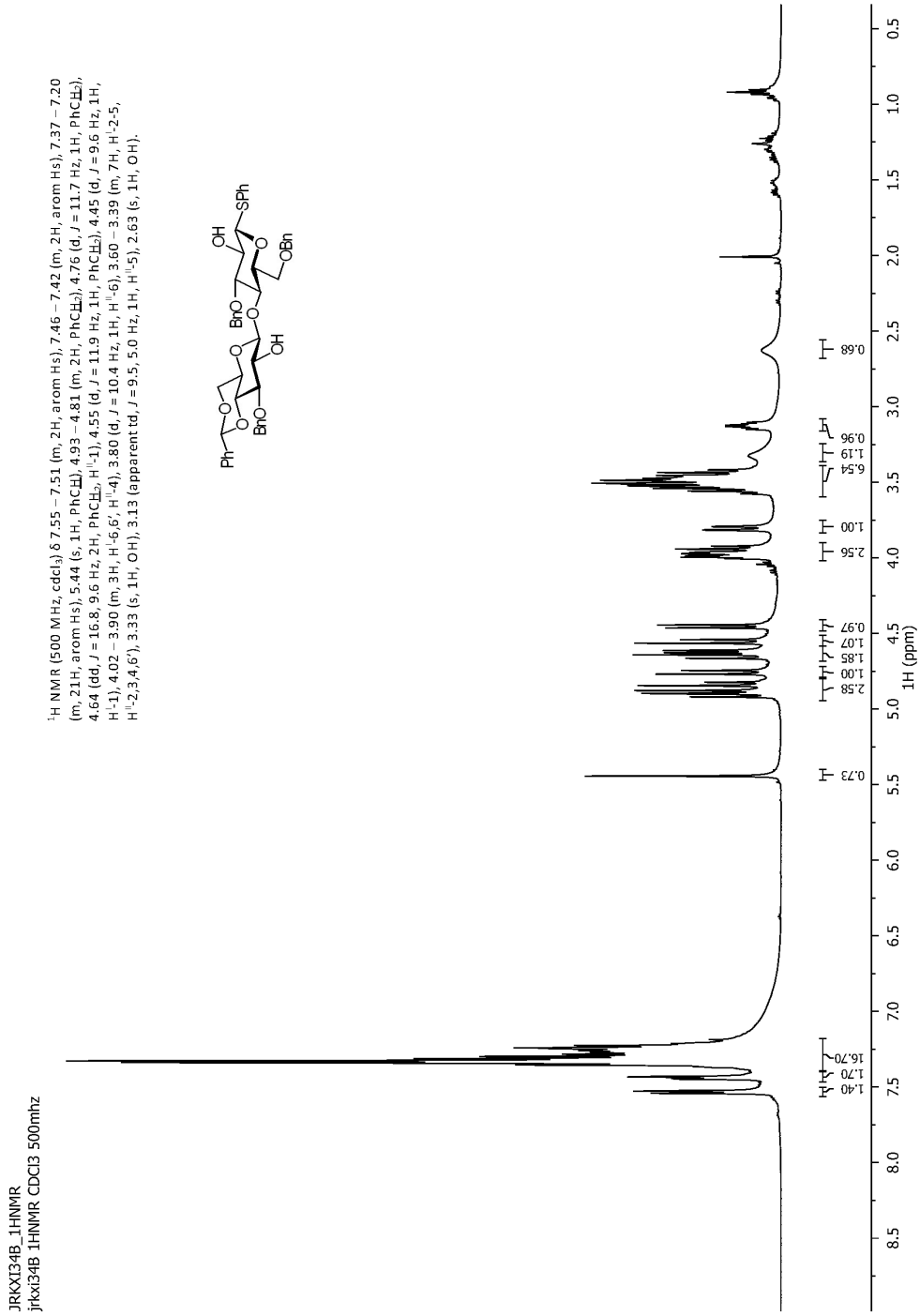


Figure 86: Compound 3.9 1H-NMR

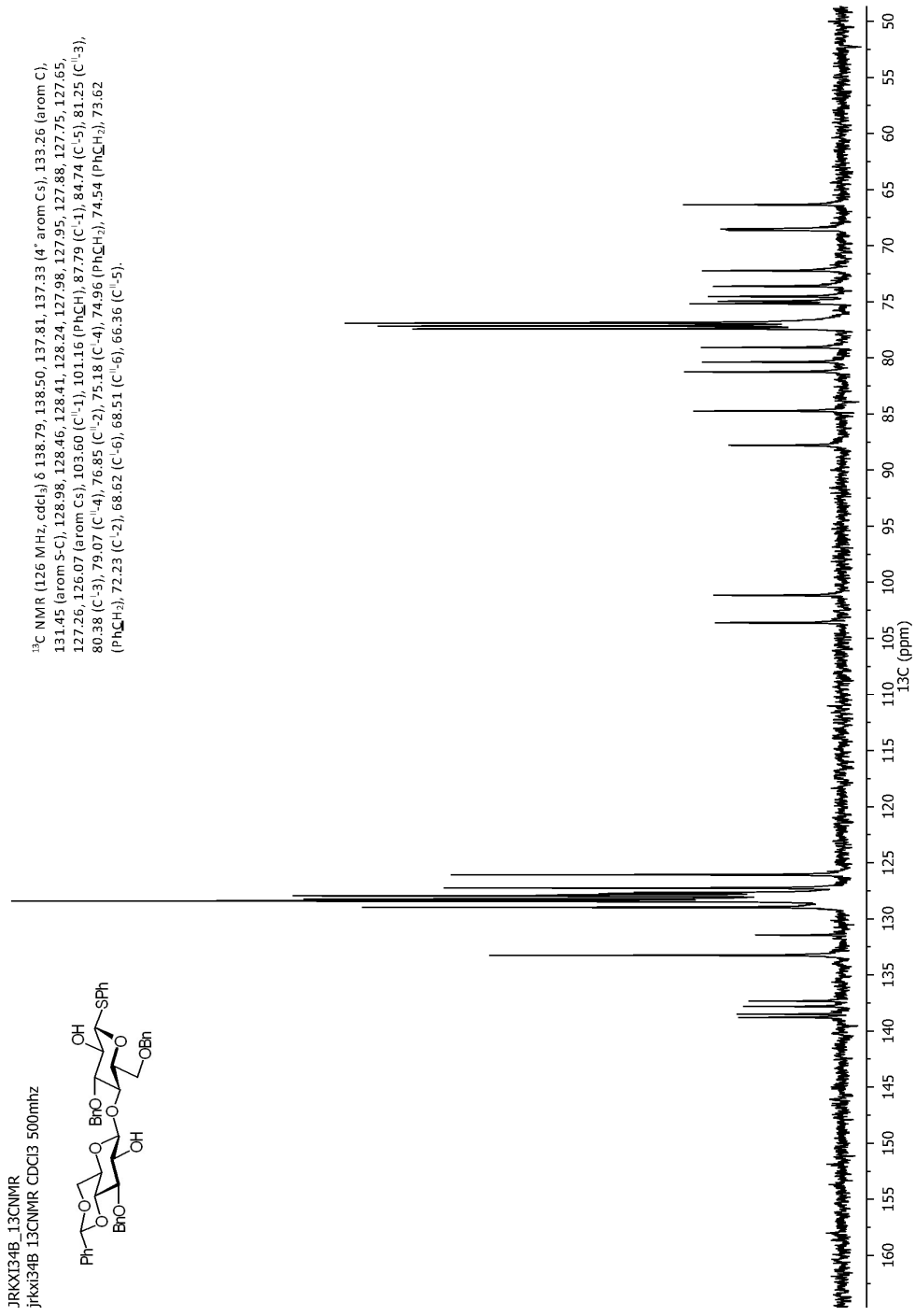


Figure 87: Compound 3.9 13C-NMR

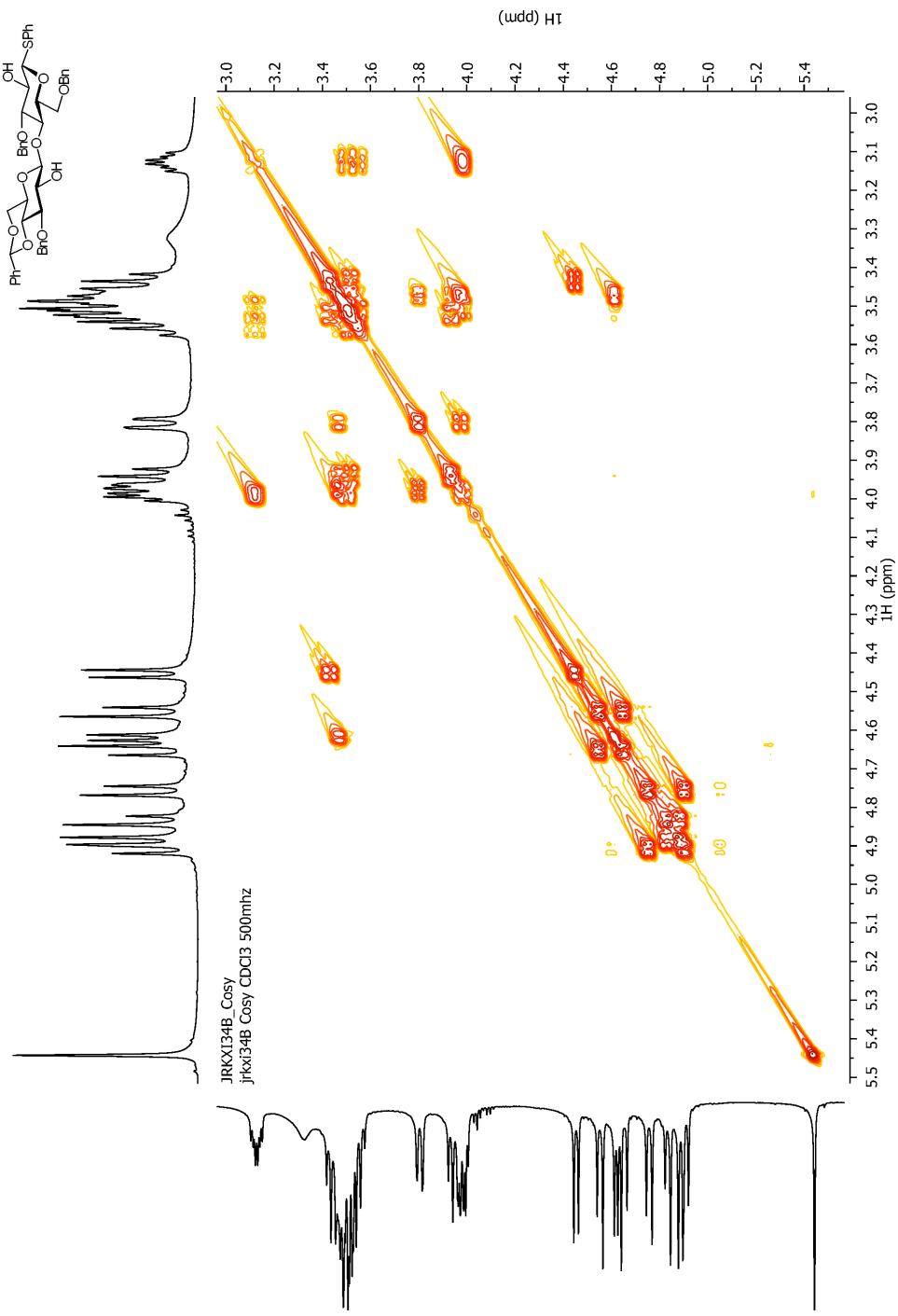


Figure 88: Compound 3.9 COSY

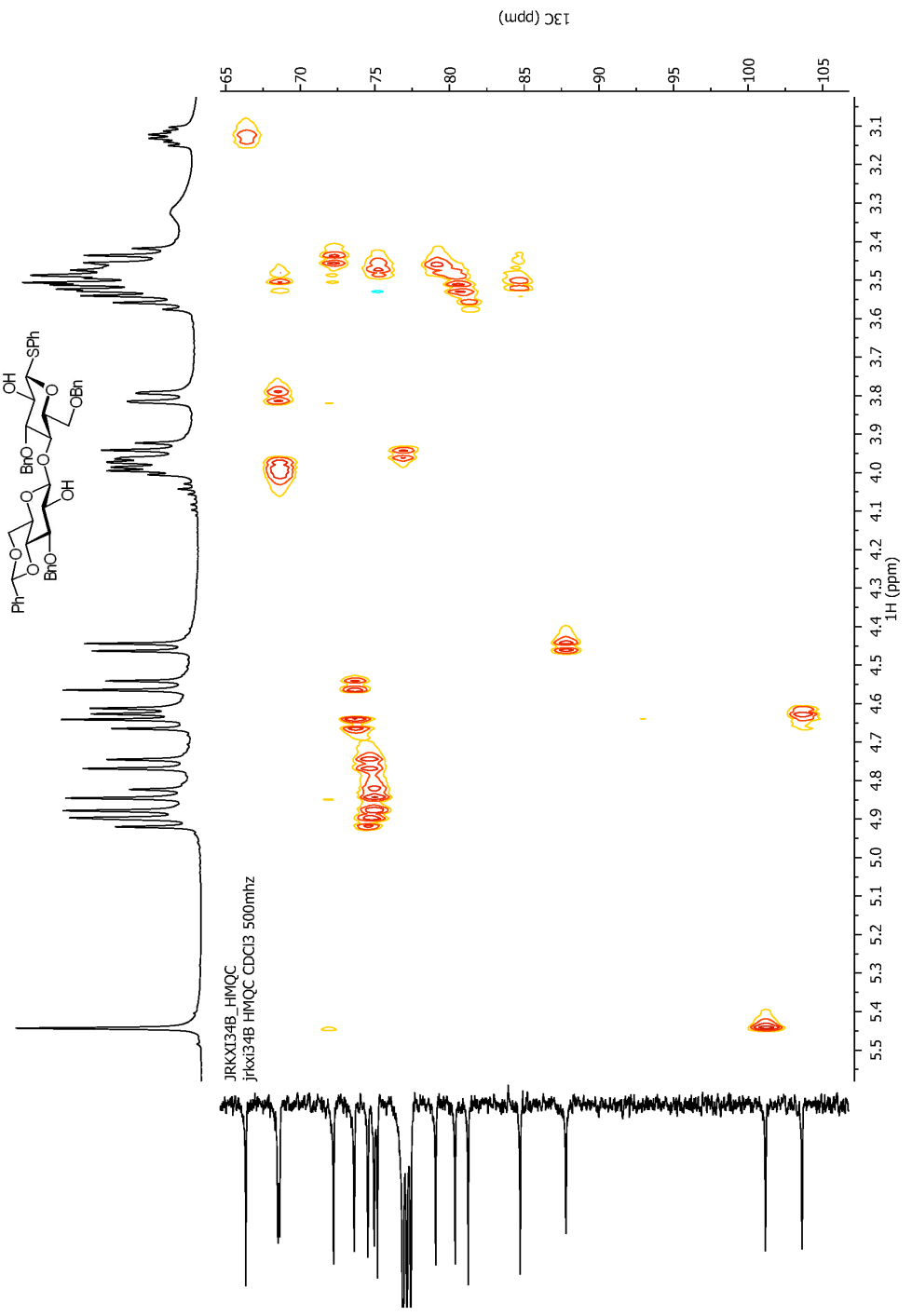


Figure 89: Compound 3.9 HMQC

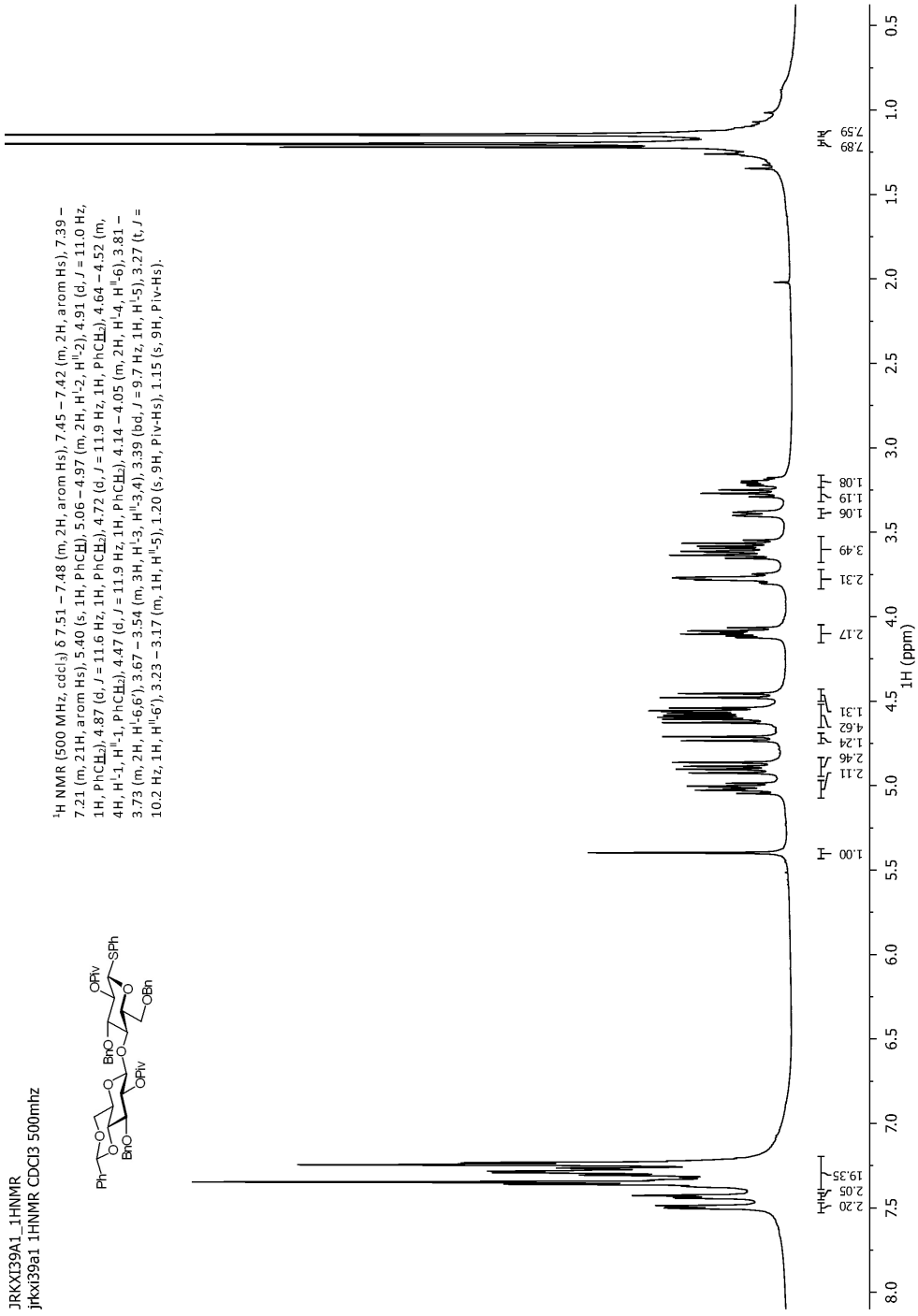


Figure 90: Compound 3.10 1H-NMR

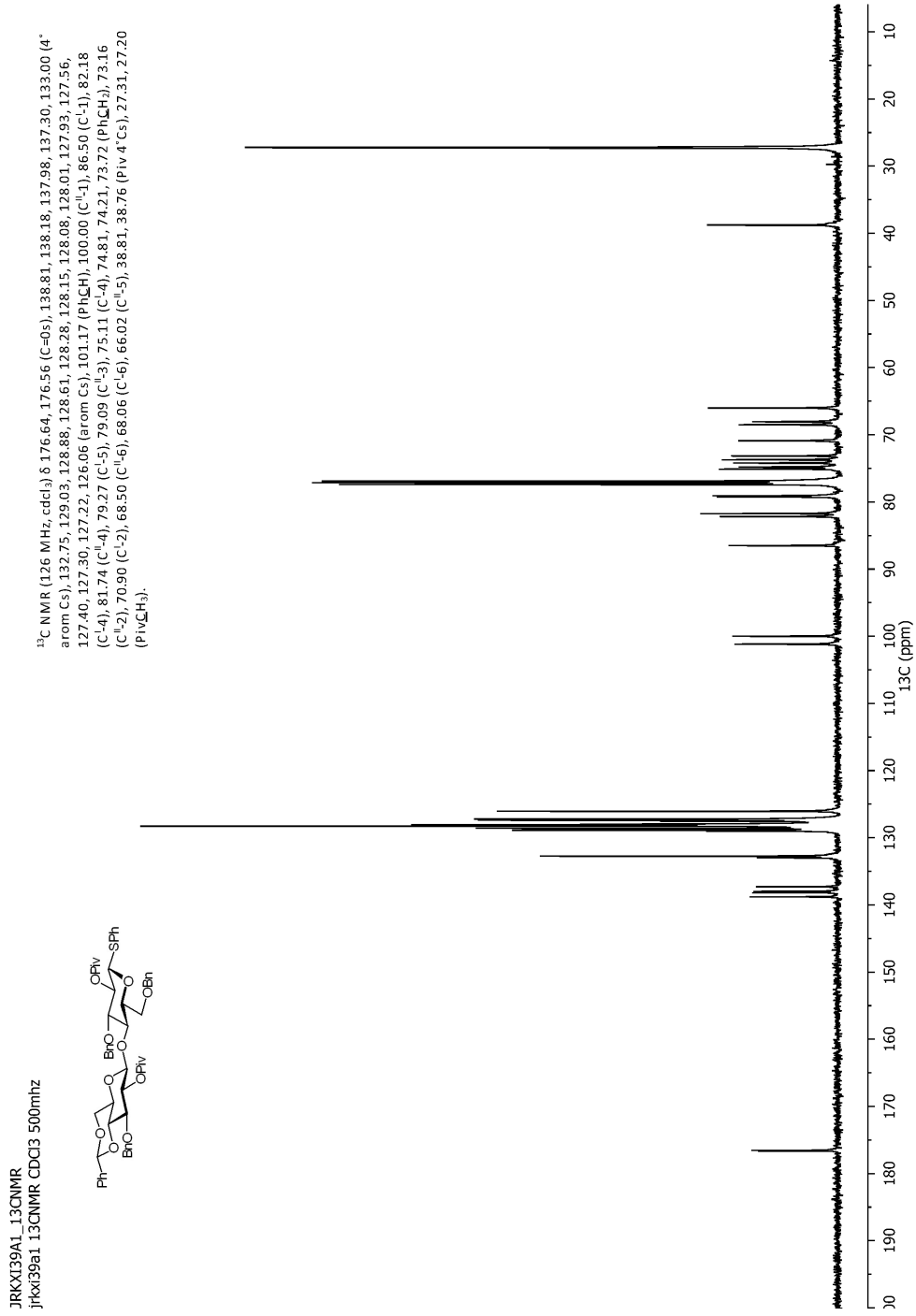


Figure 91: Compound 3.10 13C-NMR

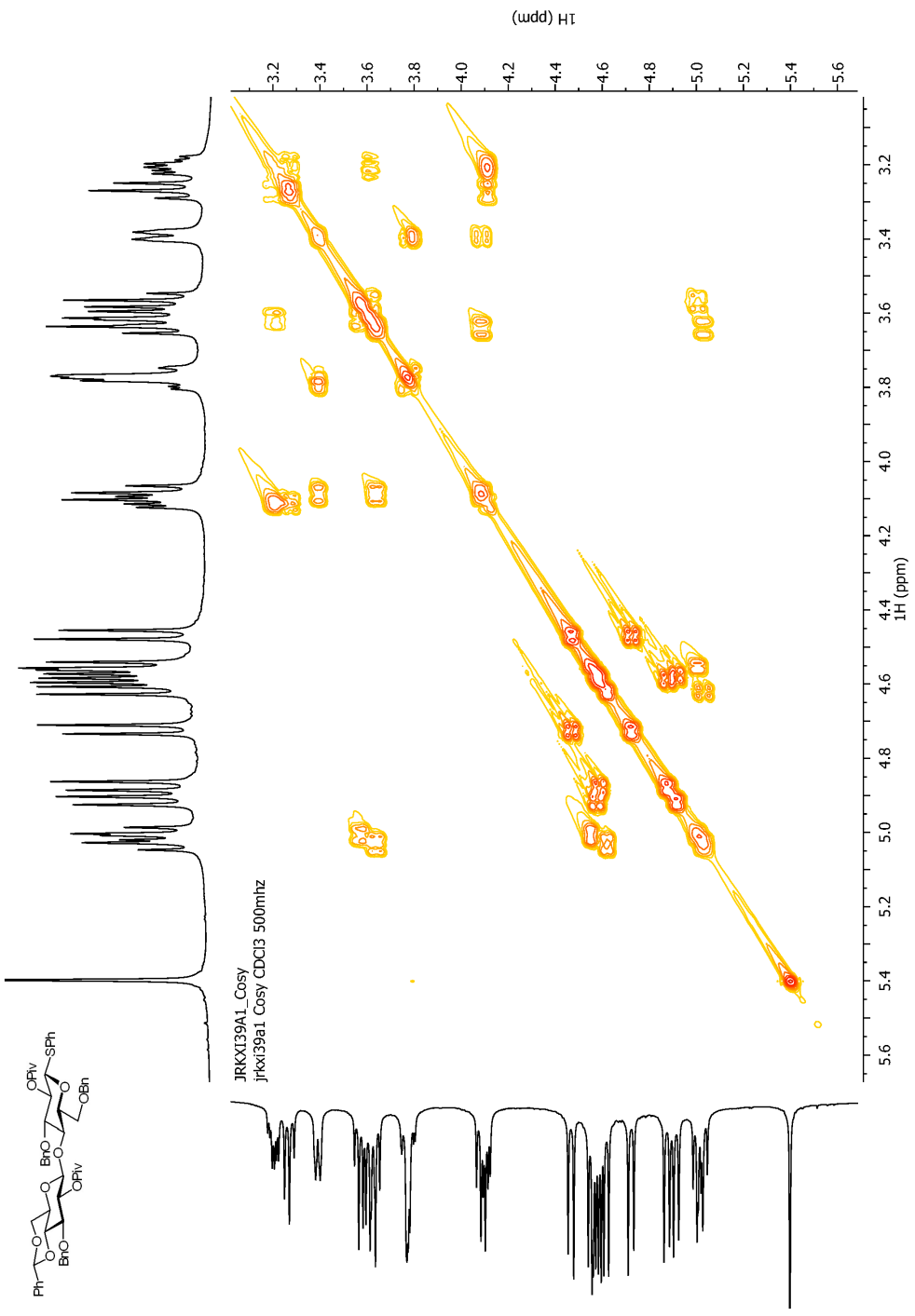


Figure 92: Compound 3.10 COSY

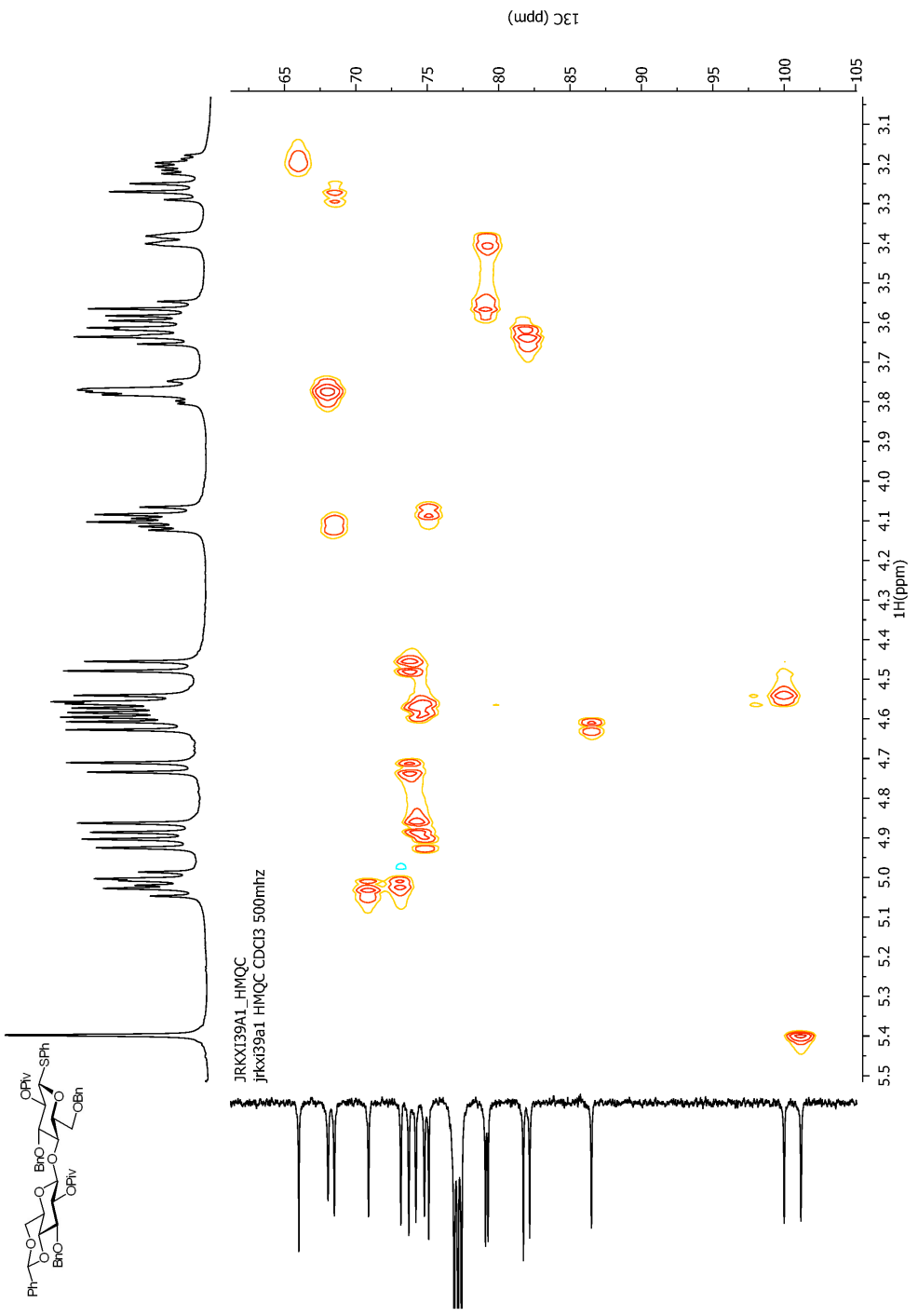


Figure 93: Compound 3.10 HMQC

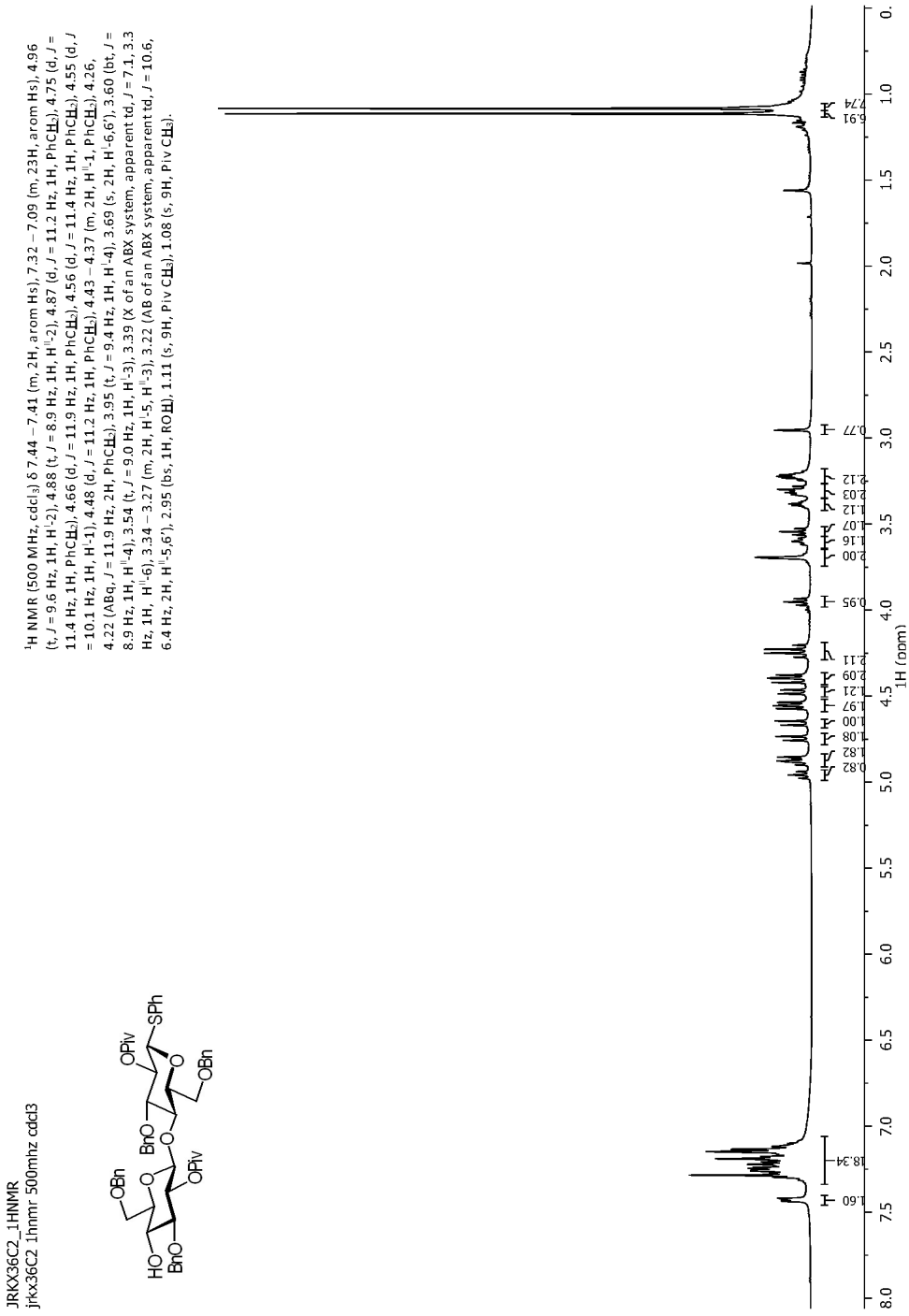


Figure 94: Compound 3.11 1H-NMR

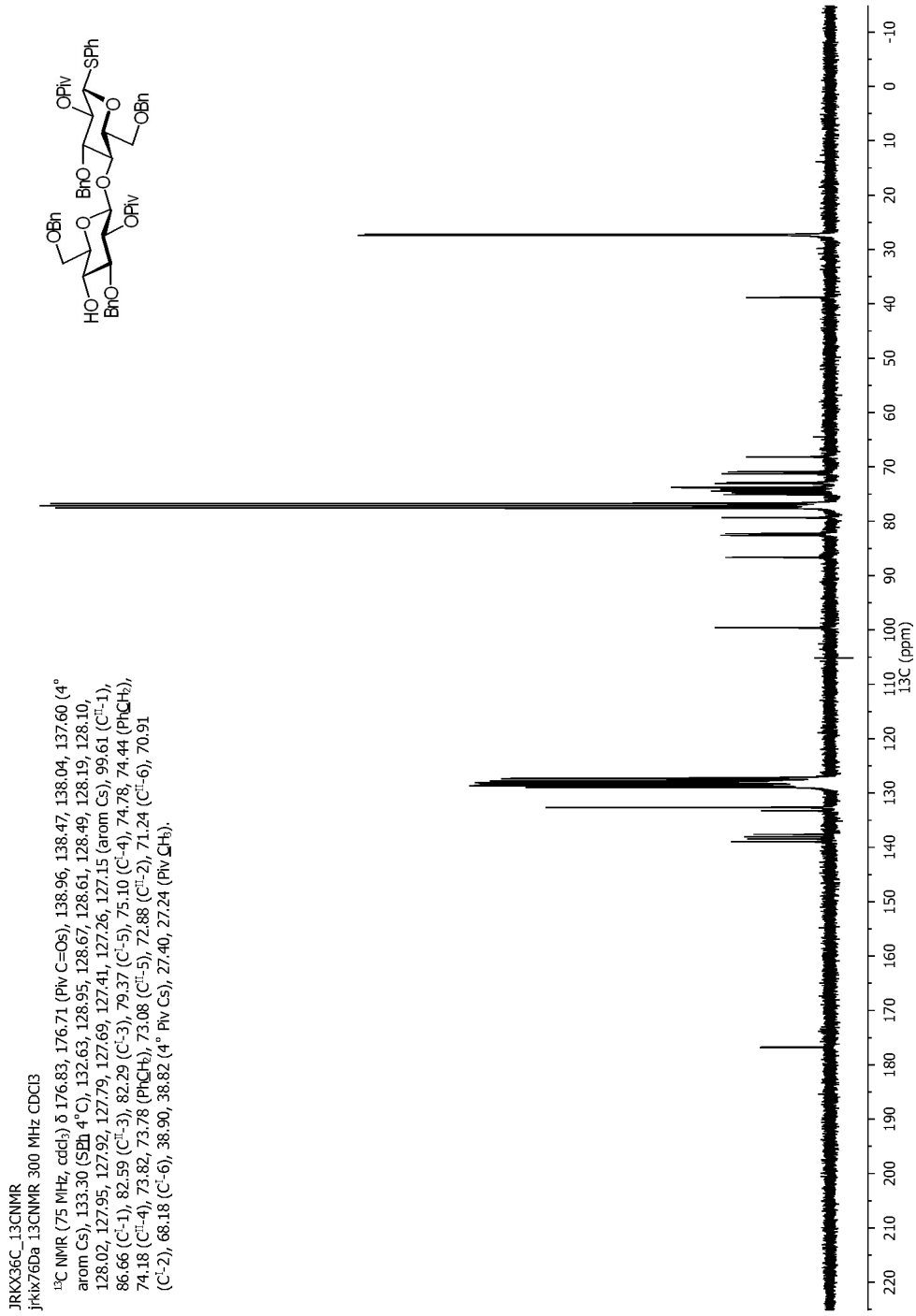


Figure 95: Compound 3.11 13C-NMR

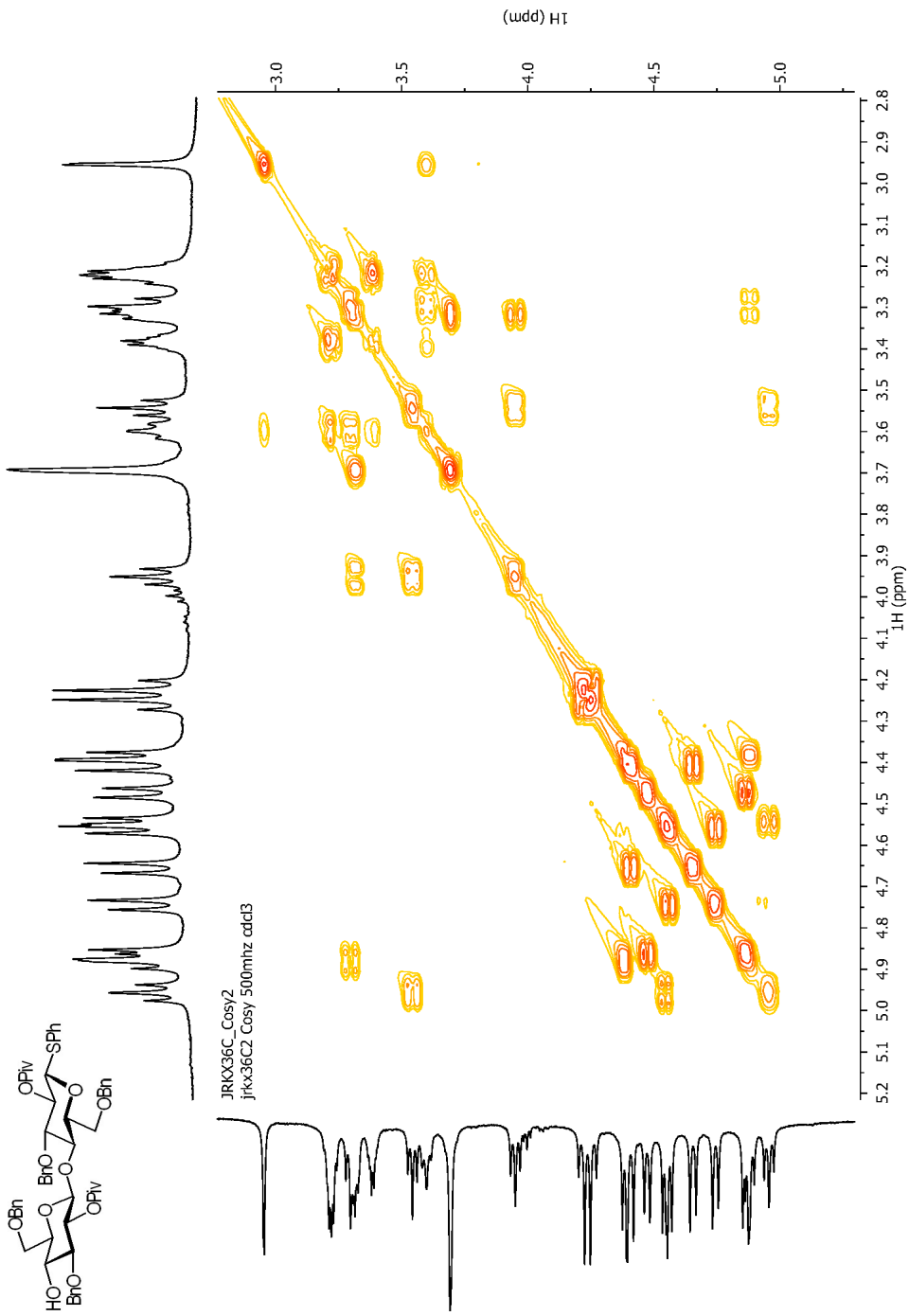


Figure 96: Compound 3.11 COSY

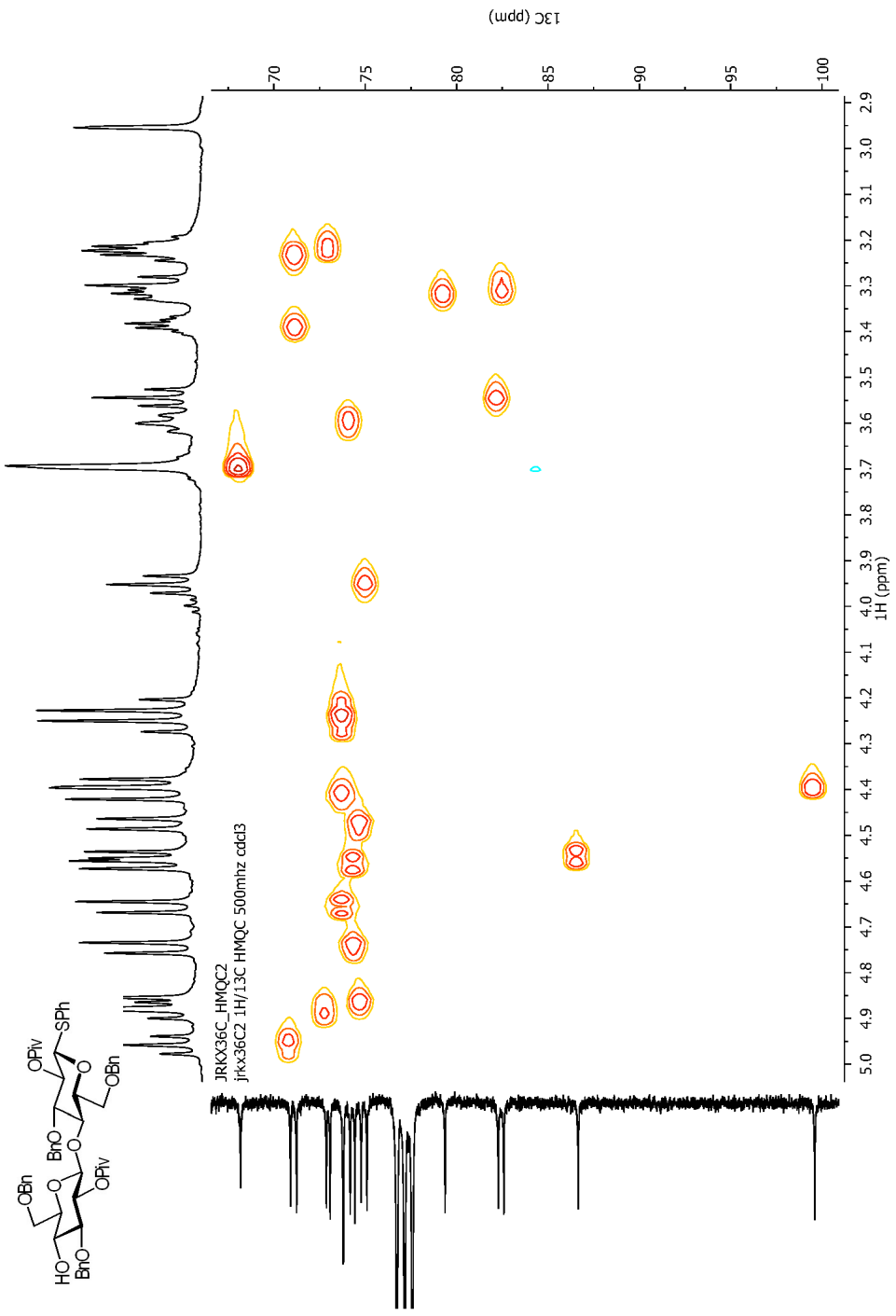


Figure 97: Compound 3.11 HMQC

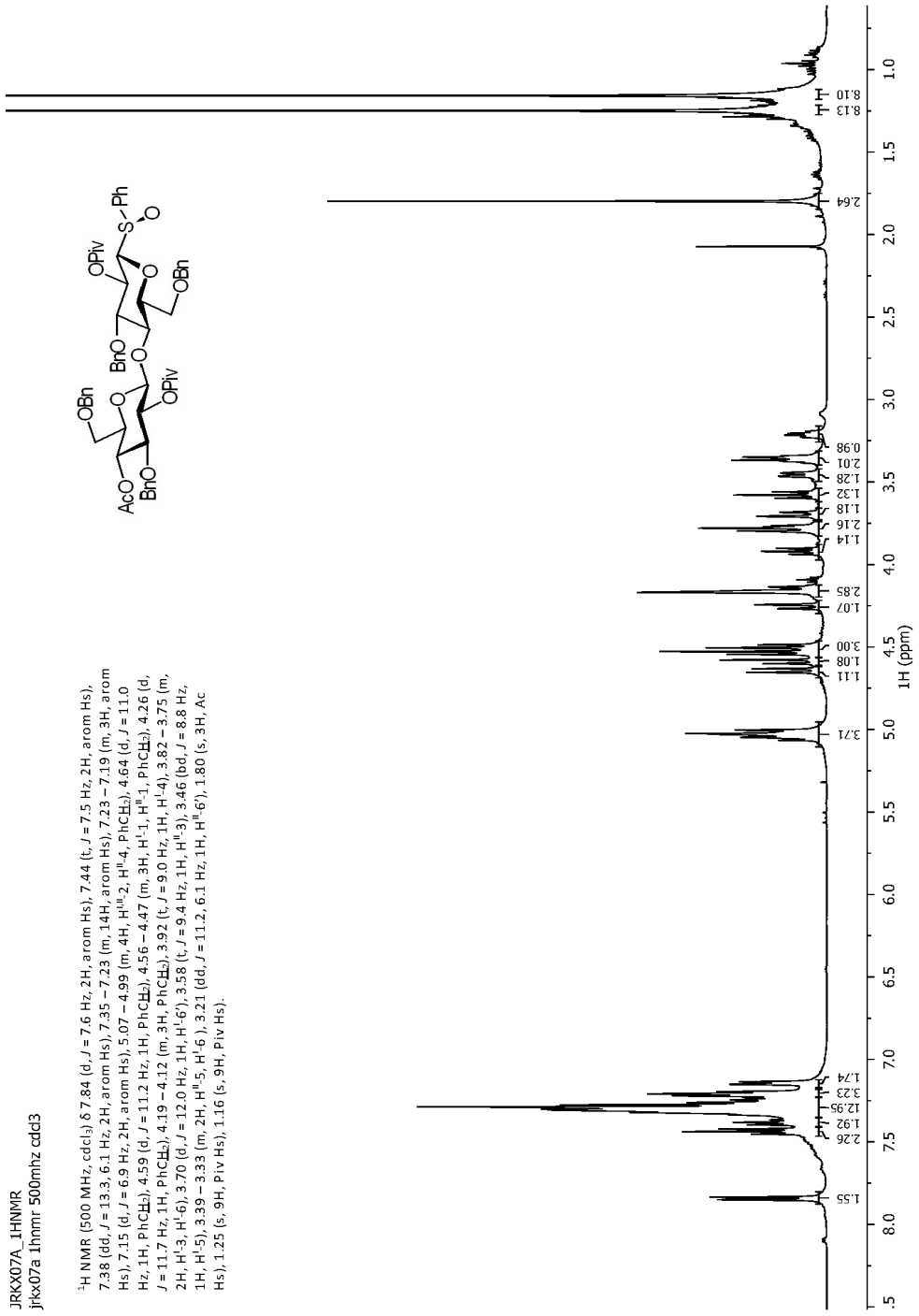


Figure 98: Compound 3.12a 1H-NMR

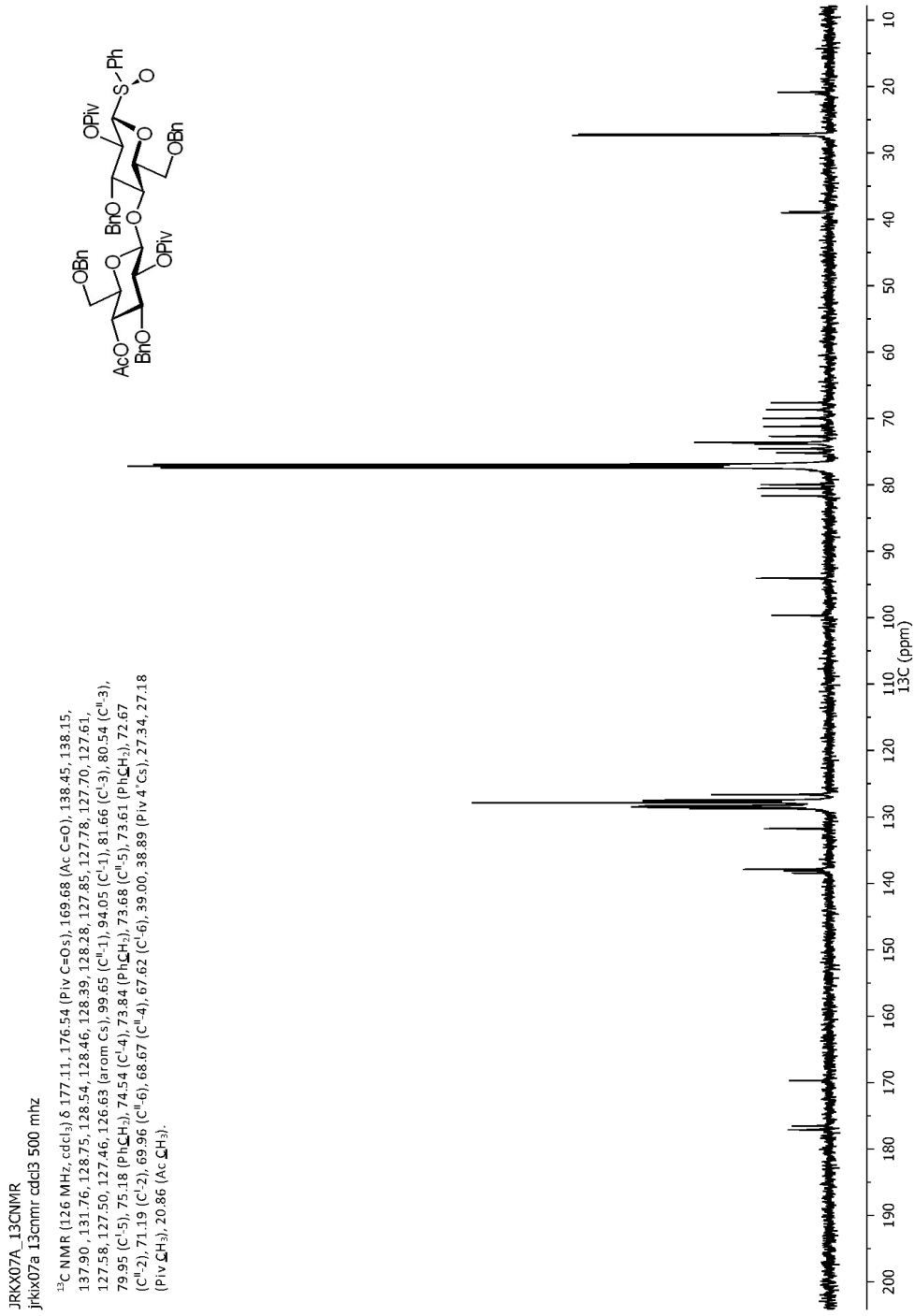


Figure 99: Compound 3.12a 13C-NMR

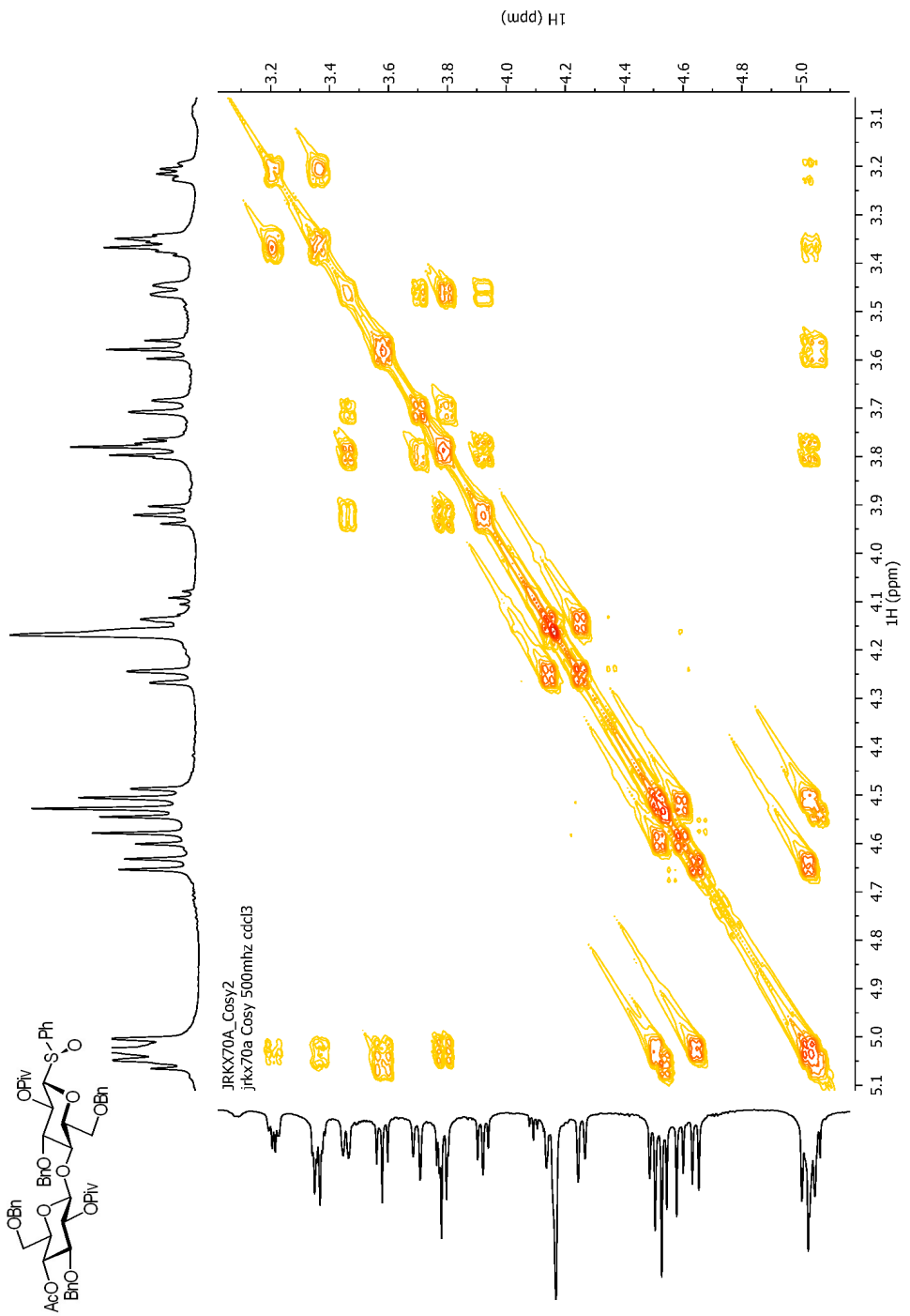


Figure 100: Compound 3.12a COSY

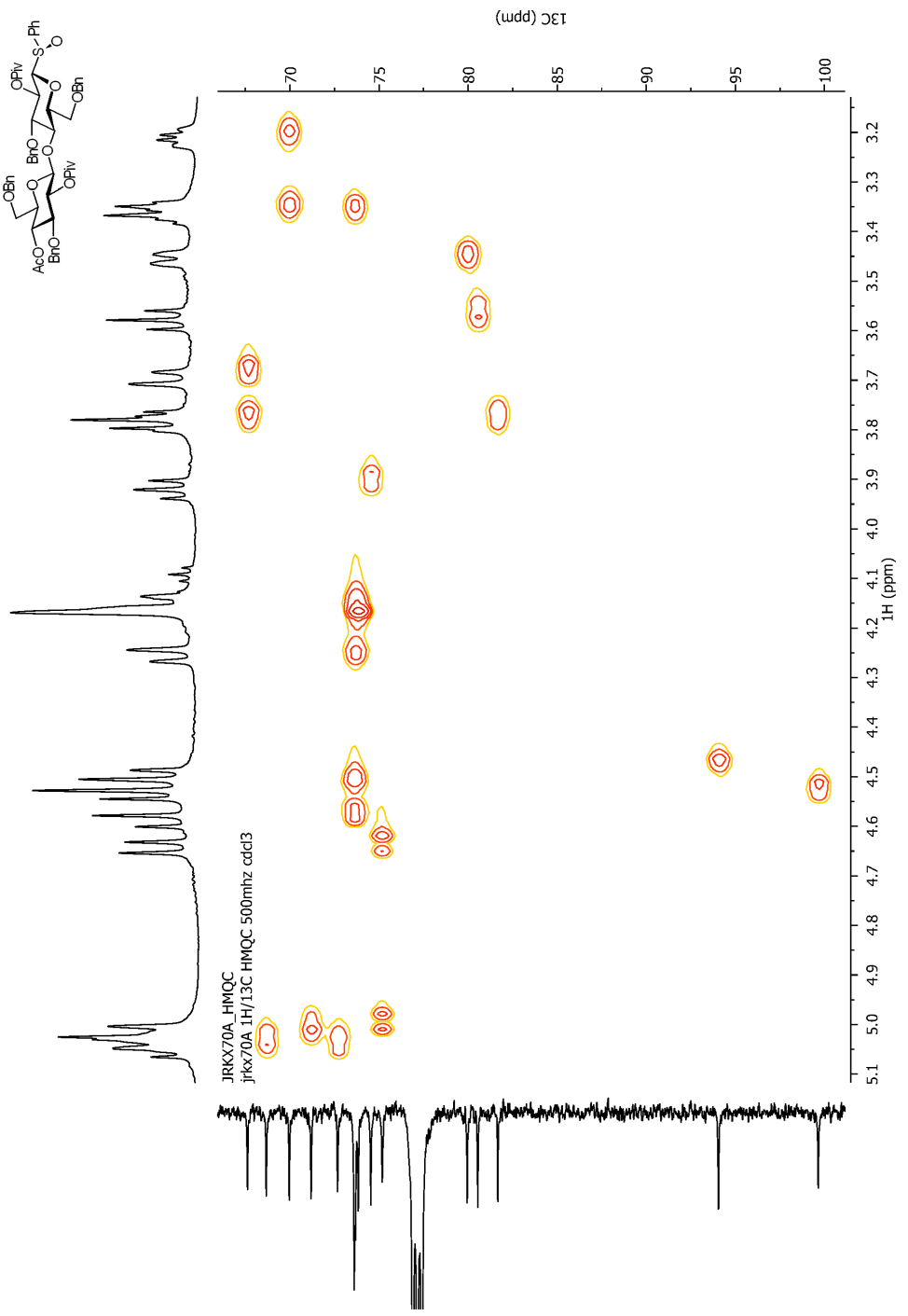


Figure 101: Compound 3.12a HMQC

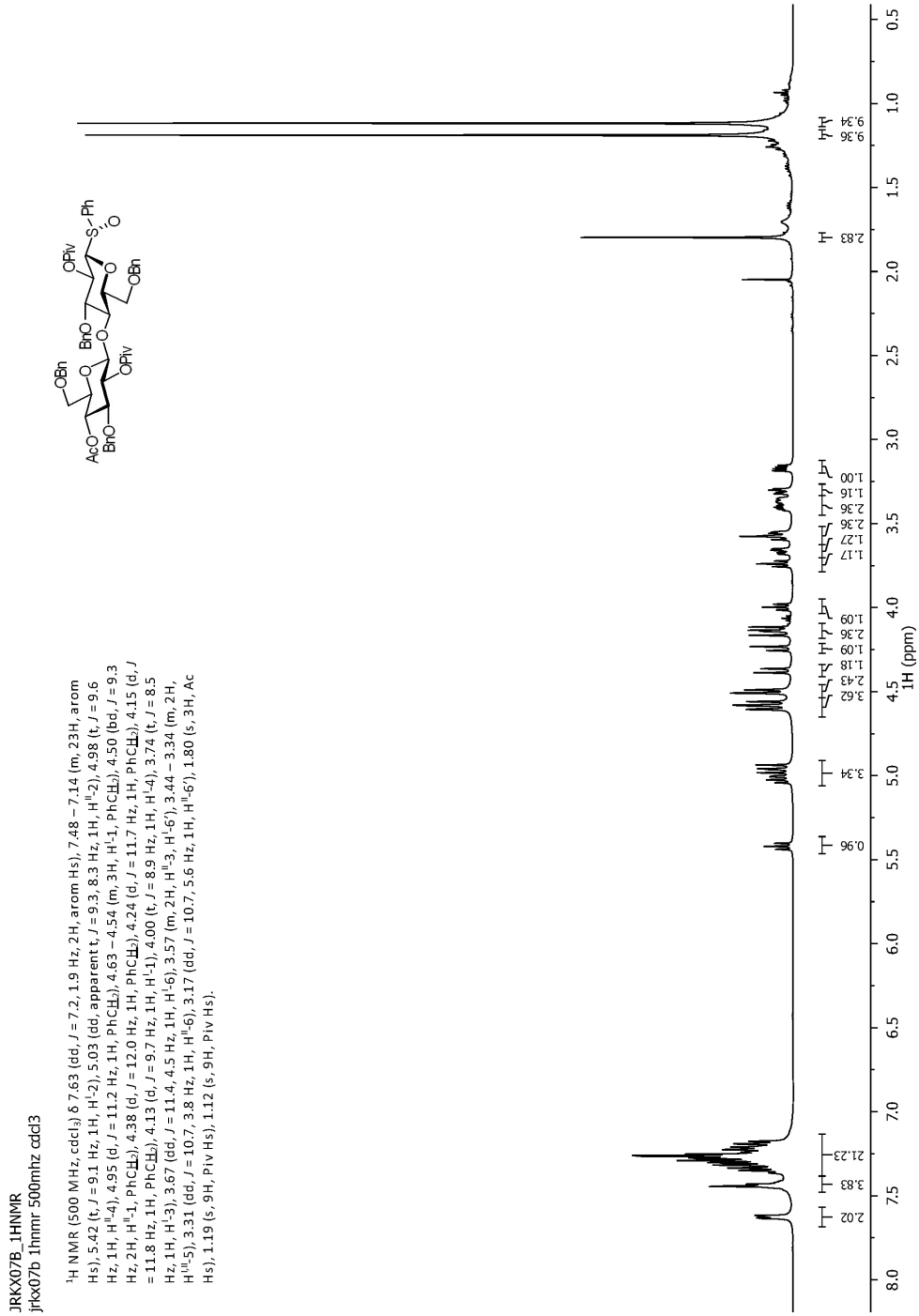


Figure 102: Compound 3.12b 1H-NMR

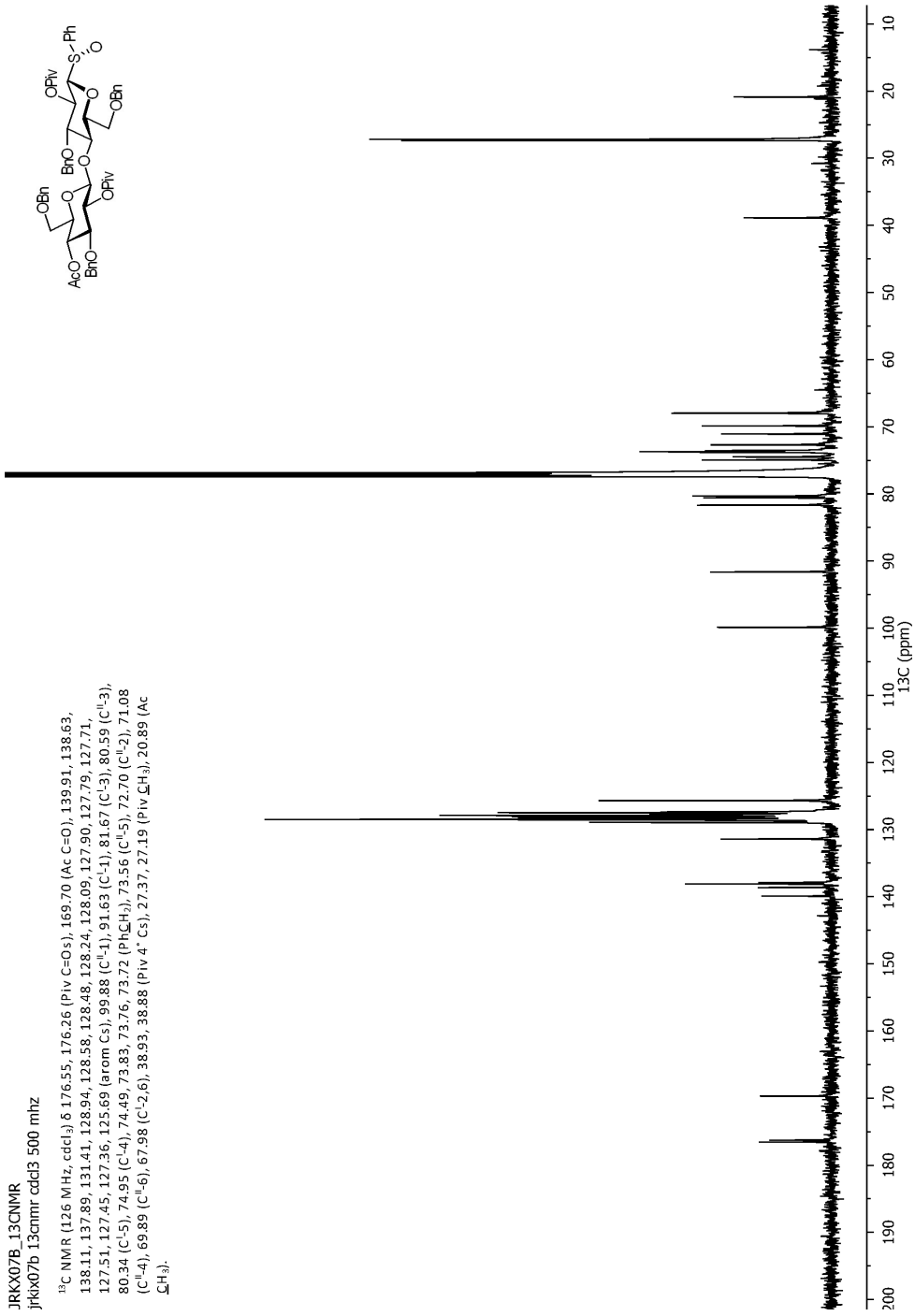


Figure 103: Compound 3.12b 13C-NMR

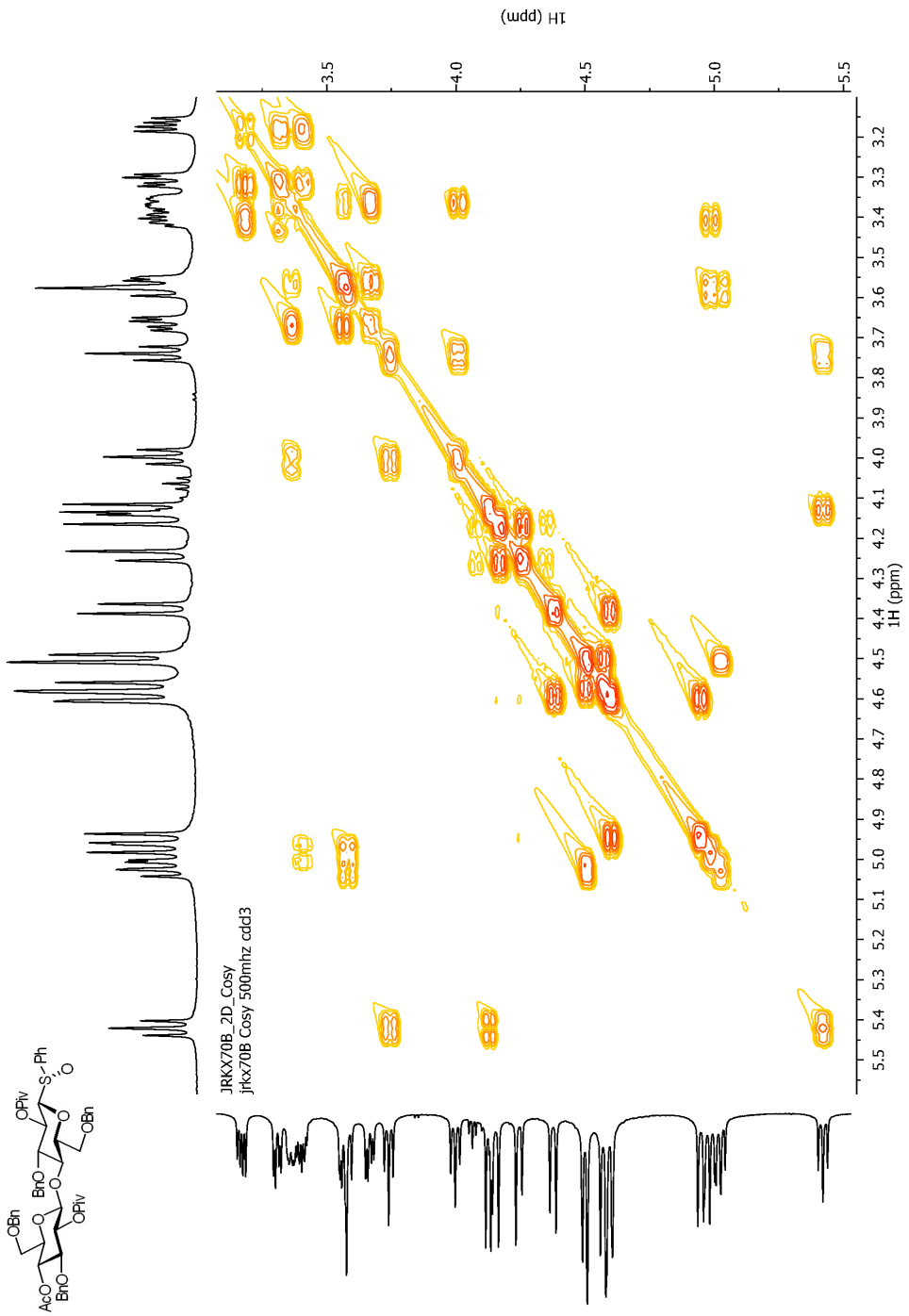


Figure 104: Compound 3.12b COSY

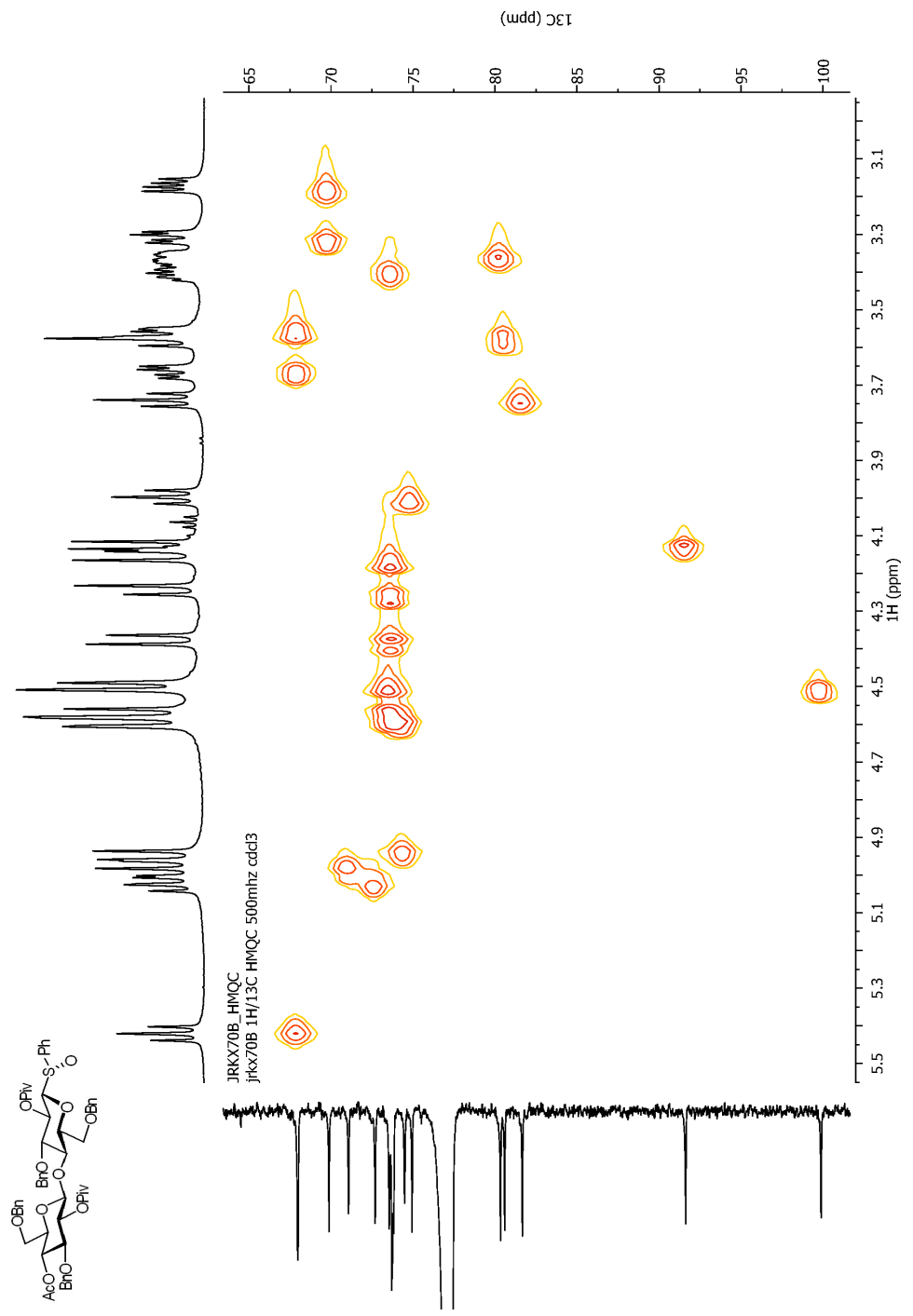


Figure 105: Compound 3.12b HMQC

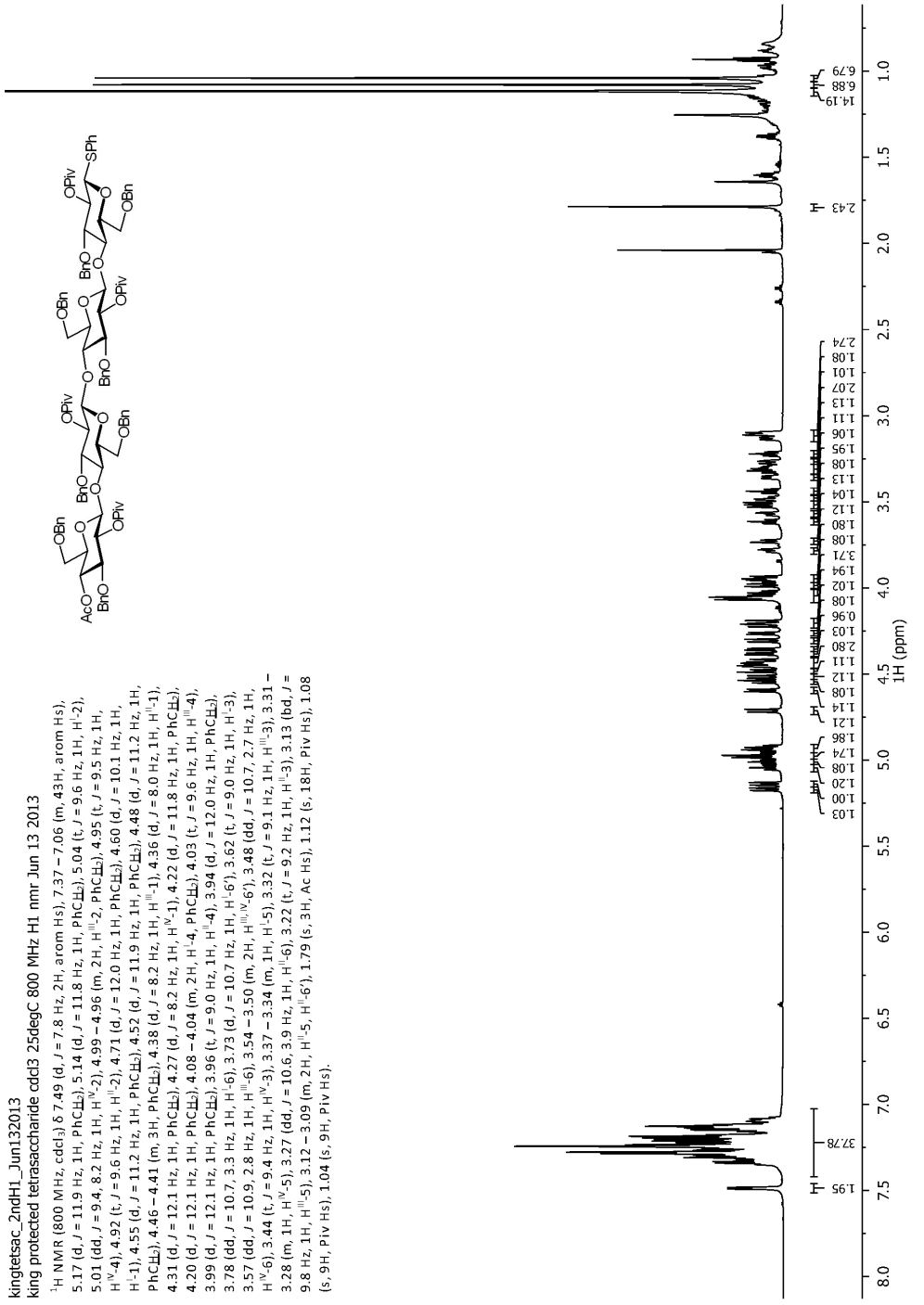


Figure 106: Compound 3.13 1H-NMR

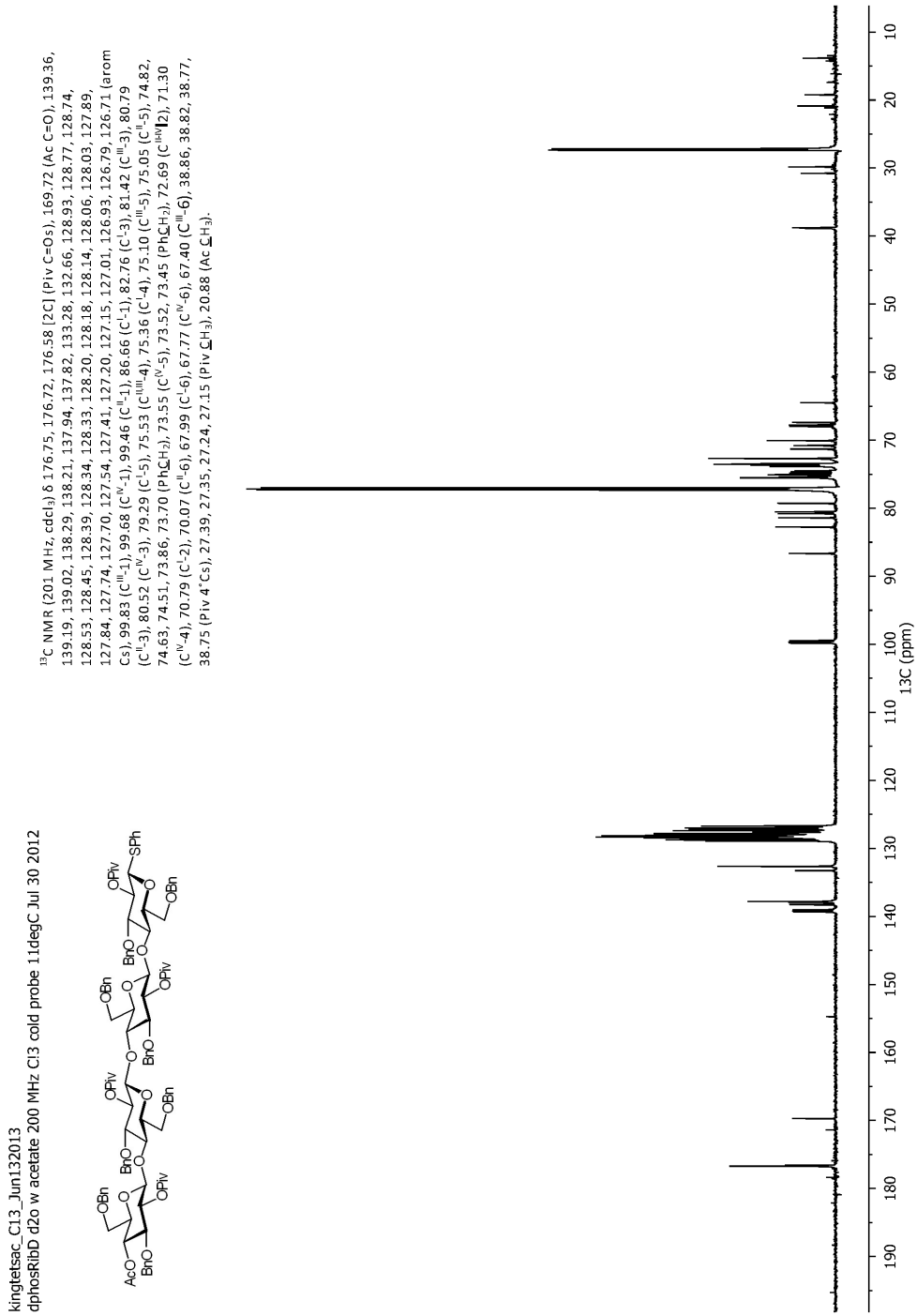


Figure 107: Compound 3.13 13C-NMR

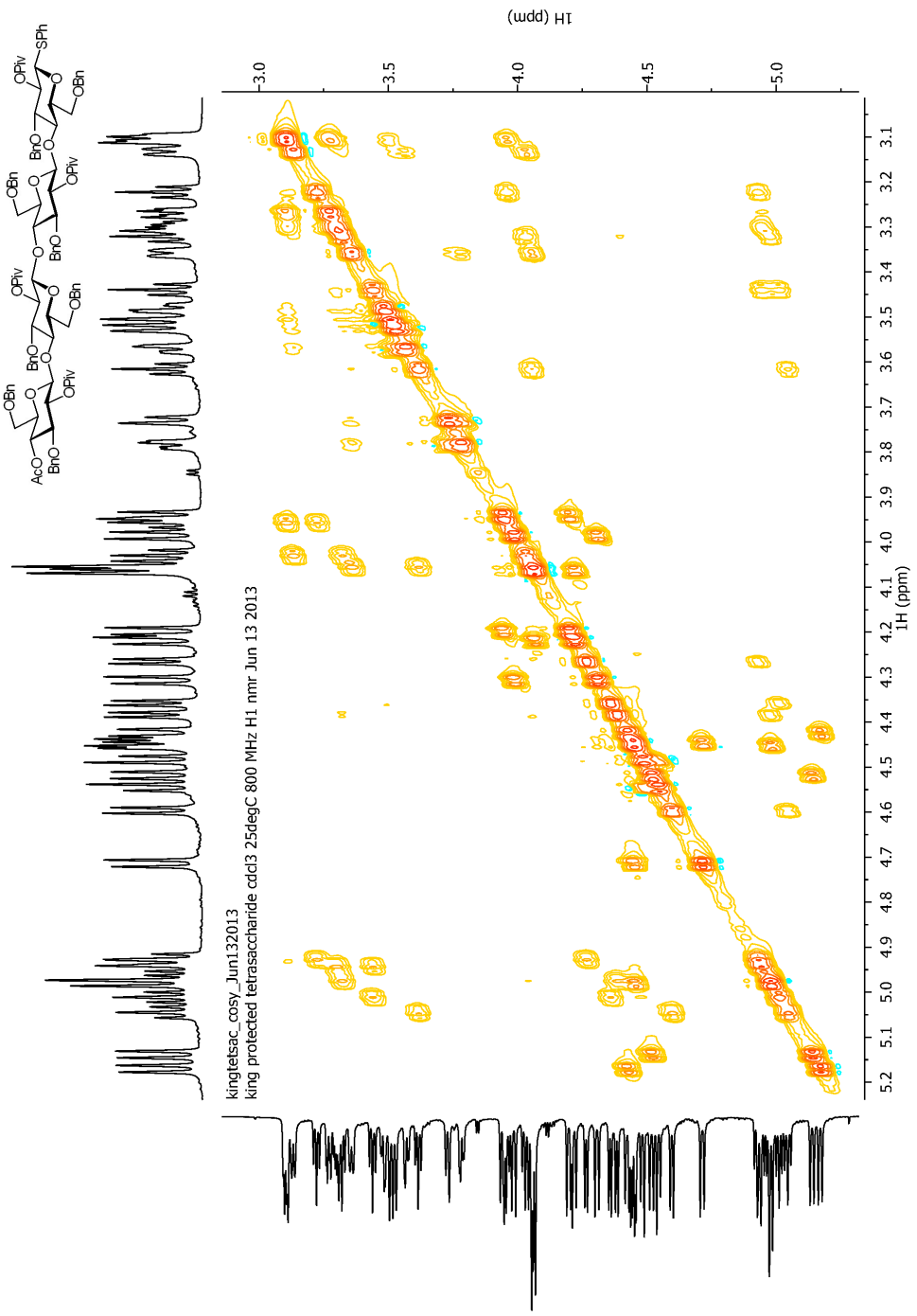


Figure 108: Compound 3.13 COSY

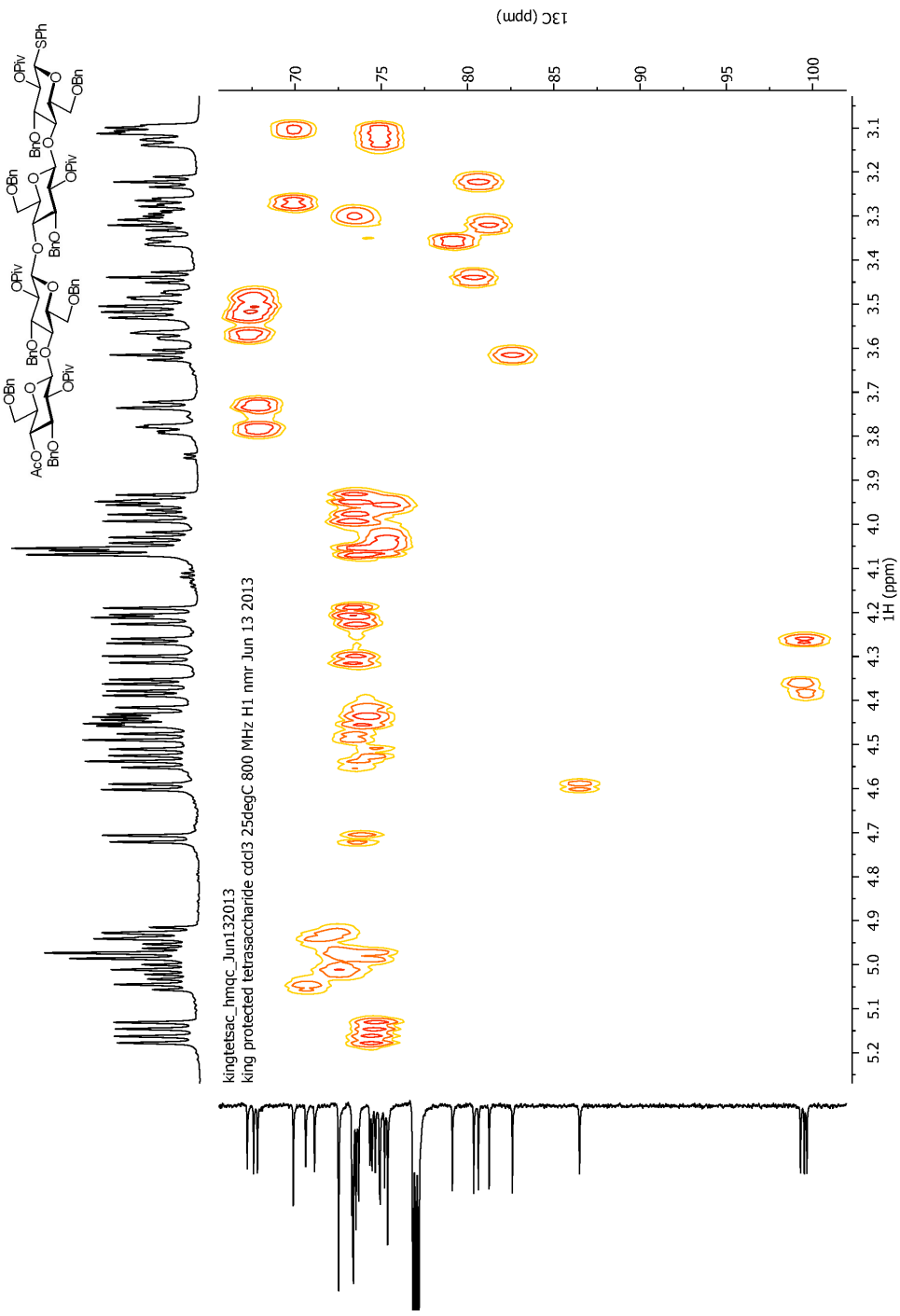


Figure 109: Compound 3.13 HMQC

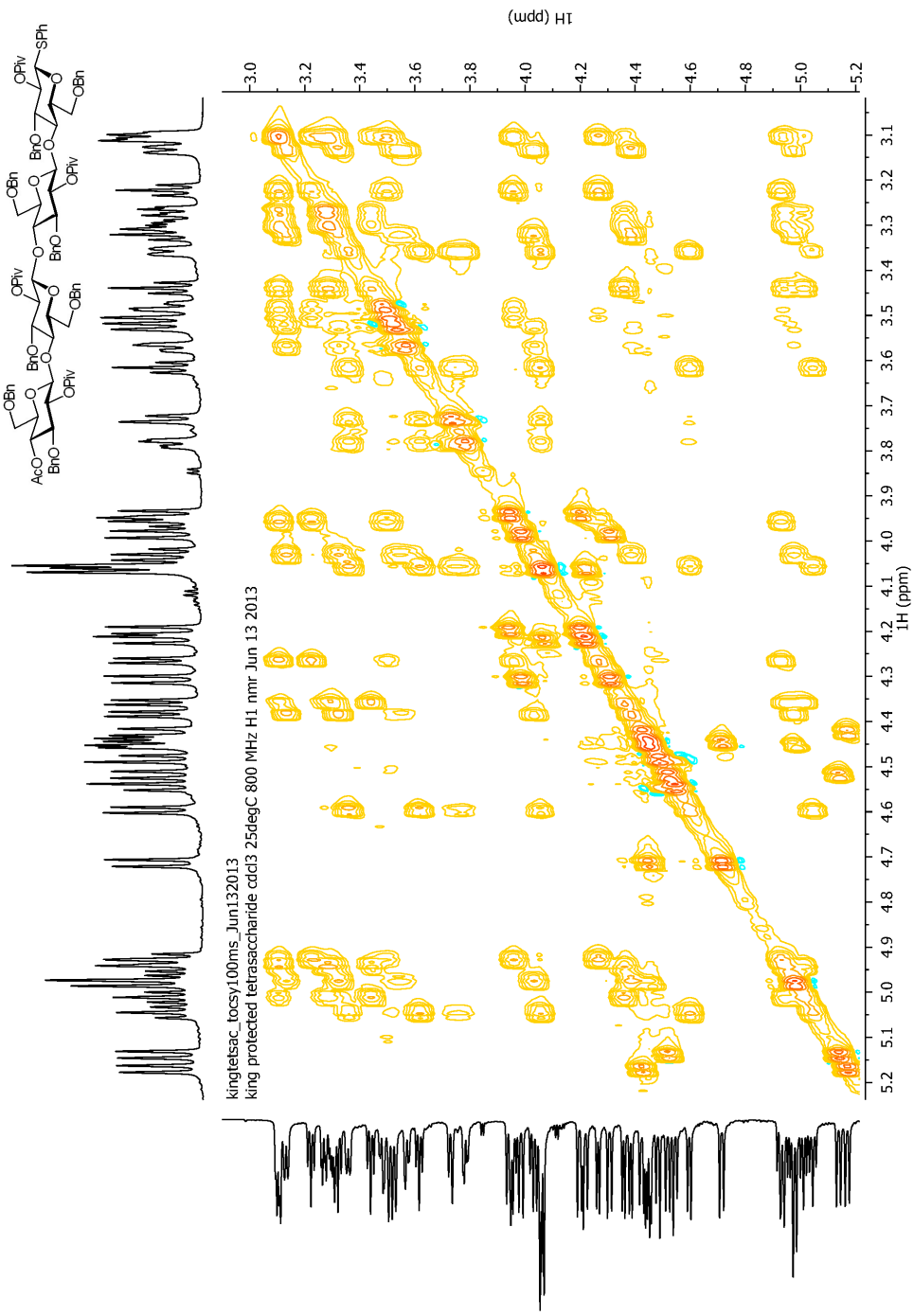


Figure 110: Compound 3.13 TOCSY

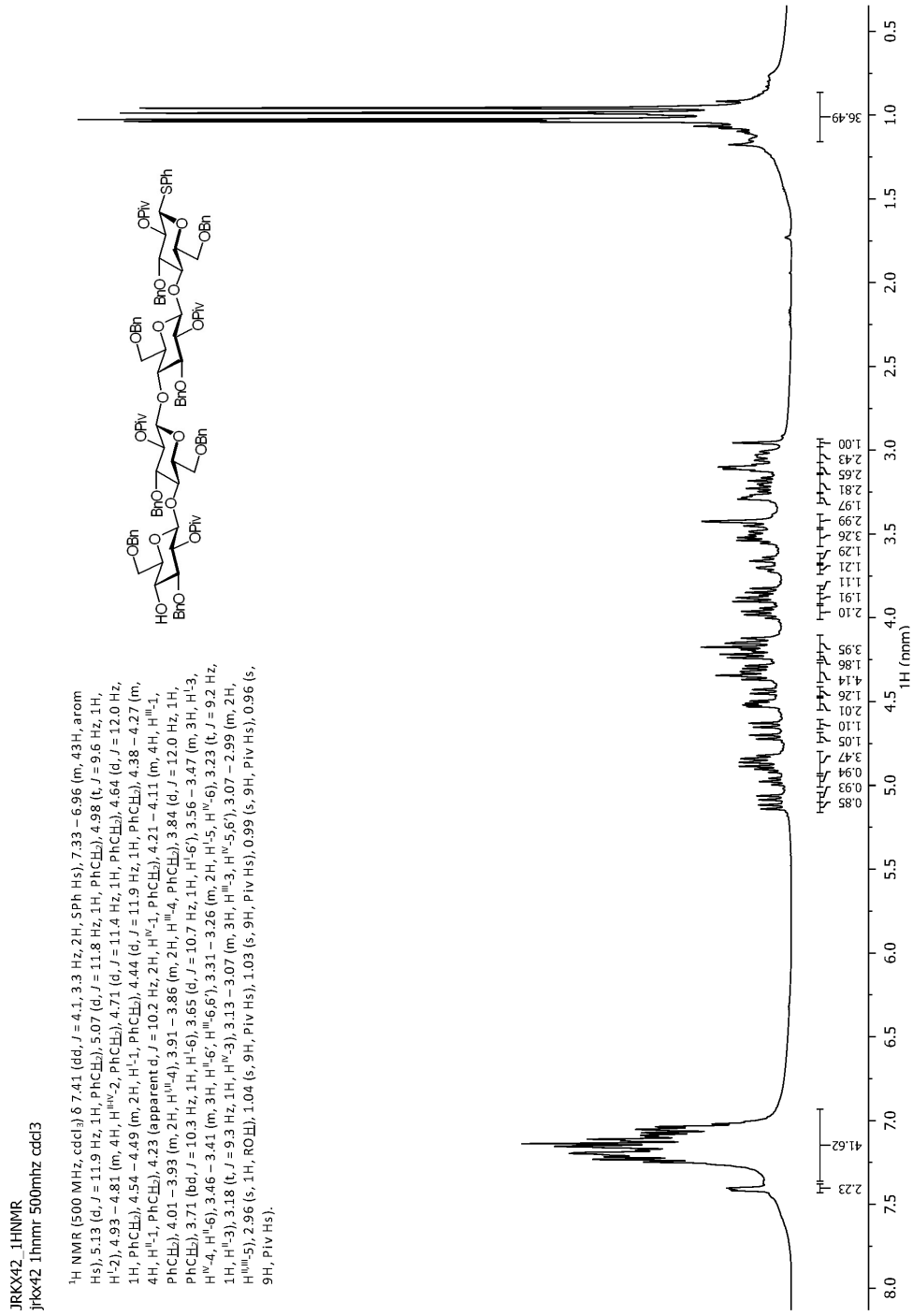


Figure 111: Compound 3.14 1H-NMR

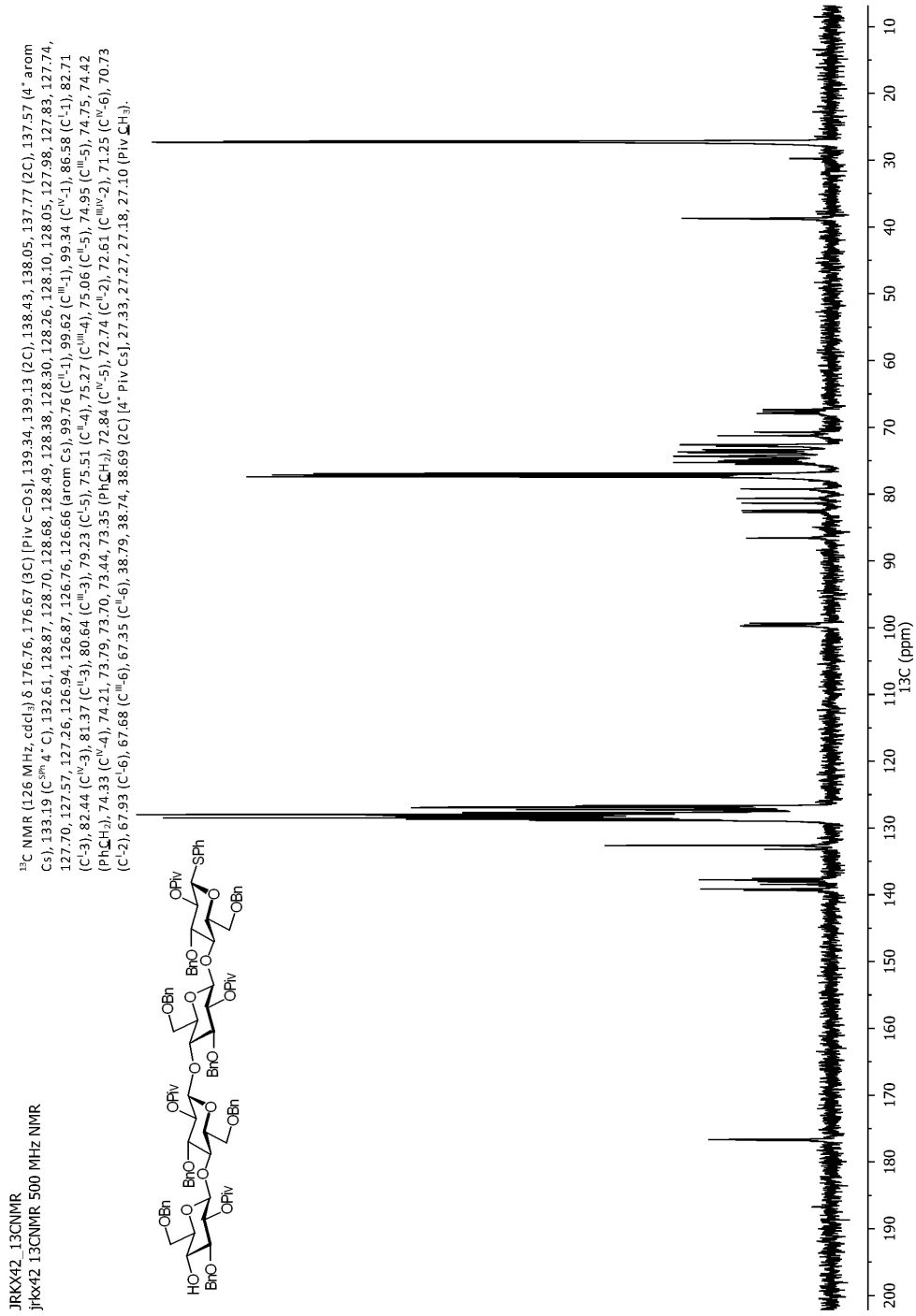


Figure 112: Compound 3.14 13C-NMR

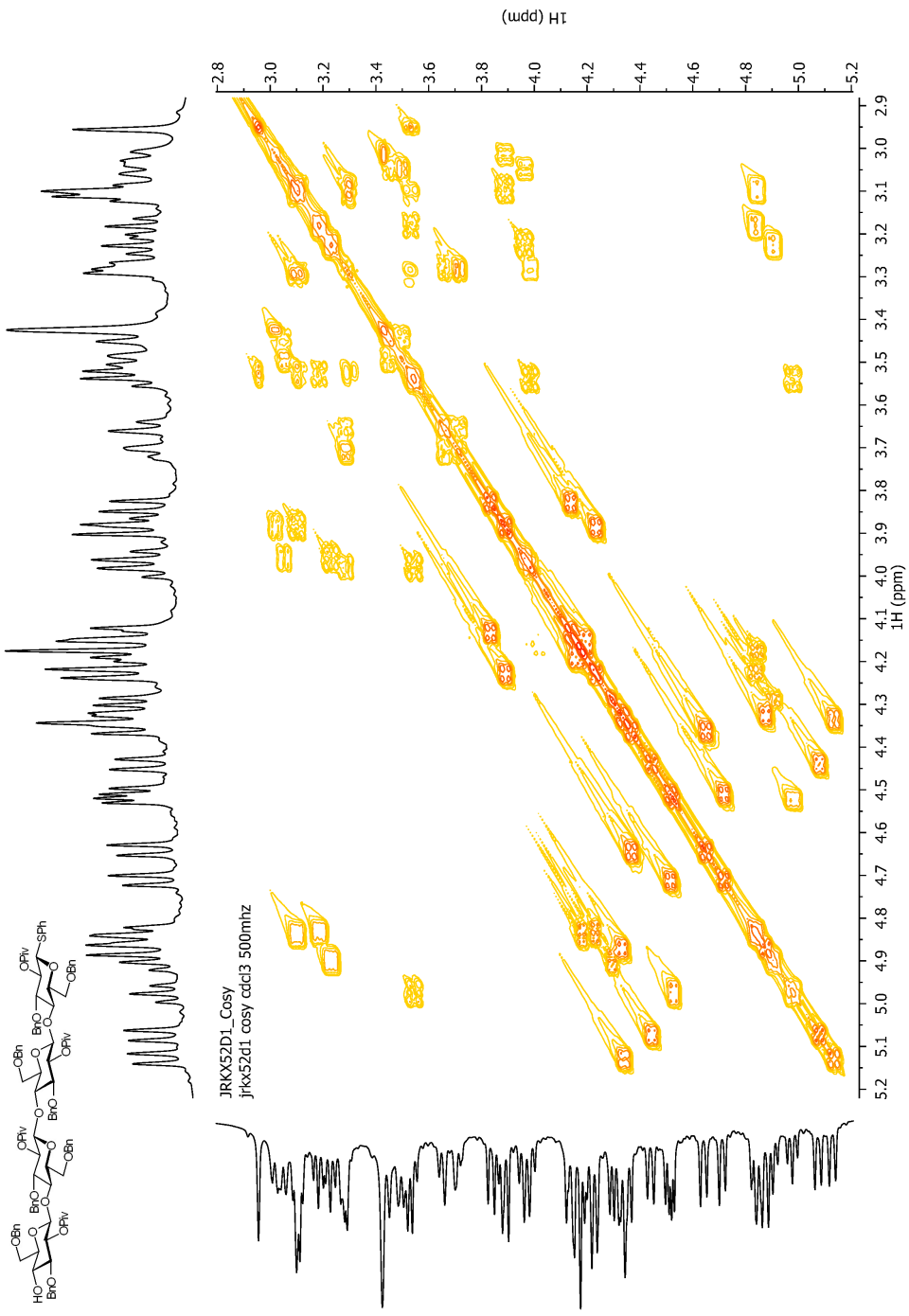


Figure 113: Compound 3.14 COSY

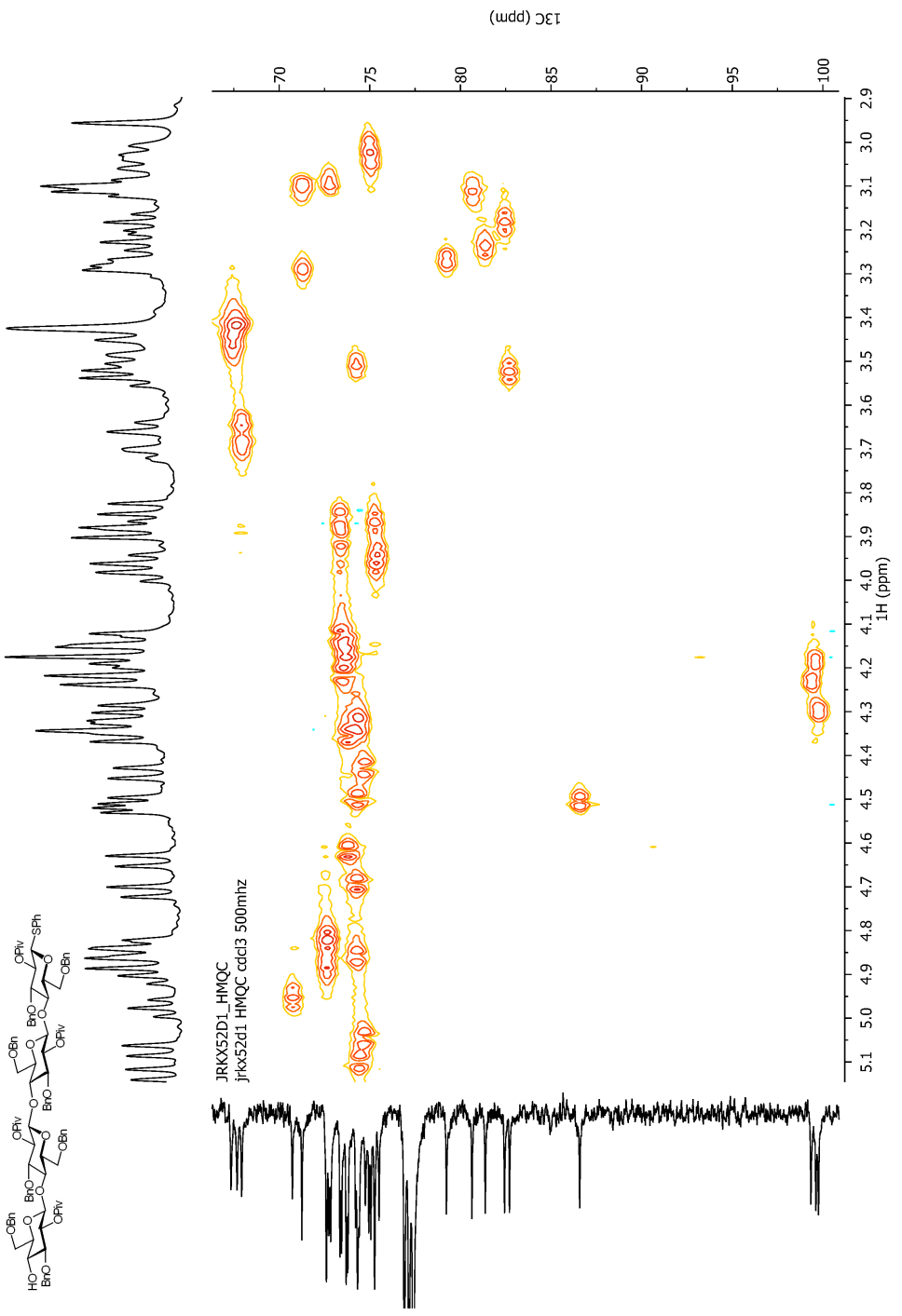


Figure 114: Compound 3.14 HMQC

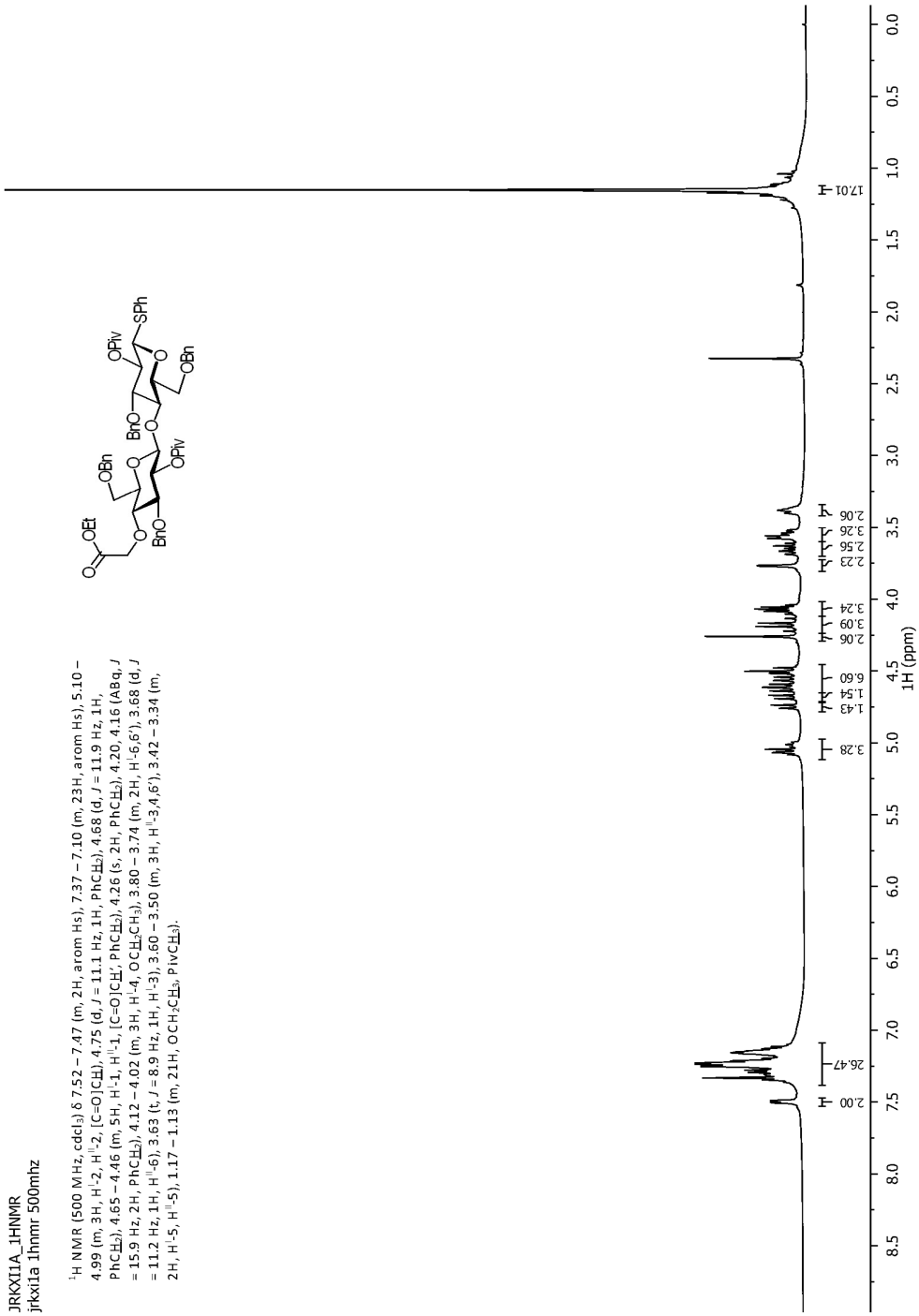


Figure 115: Compound 3.15 1H-NMR

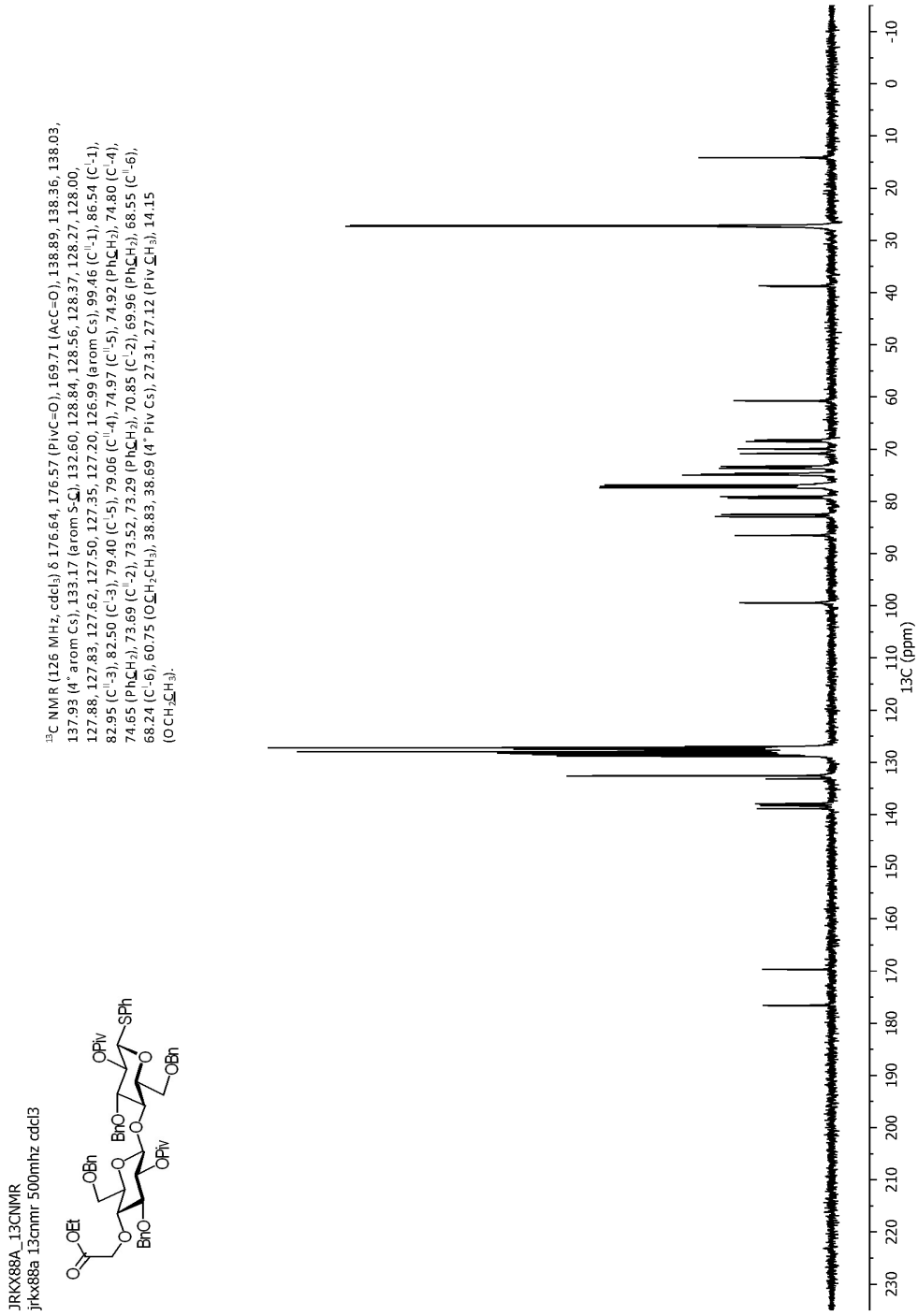


Figure 116: Compound 3.15 13C-NMR

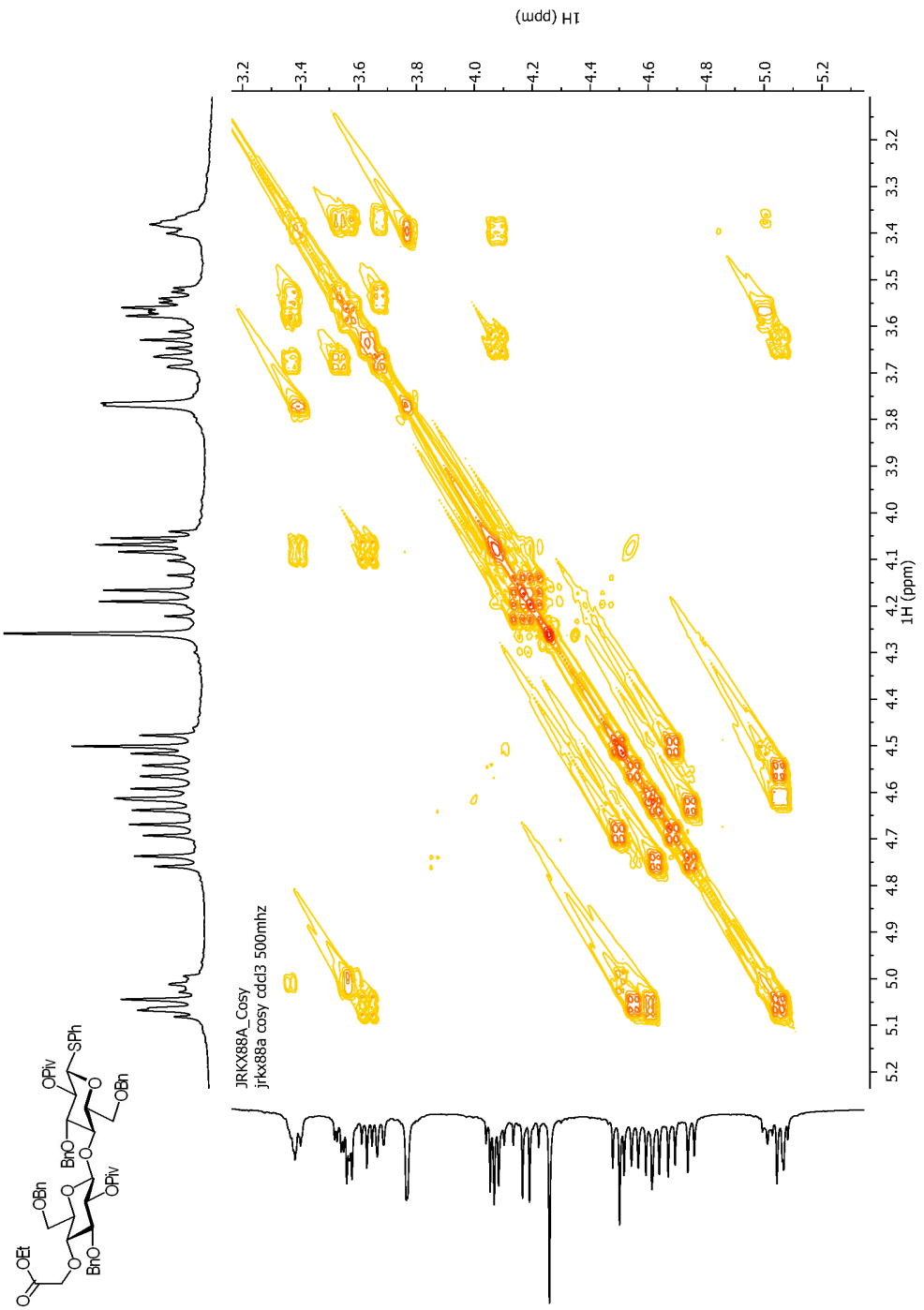


Figure 117: Compound 3.15 COSY

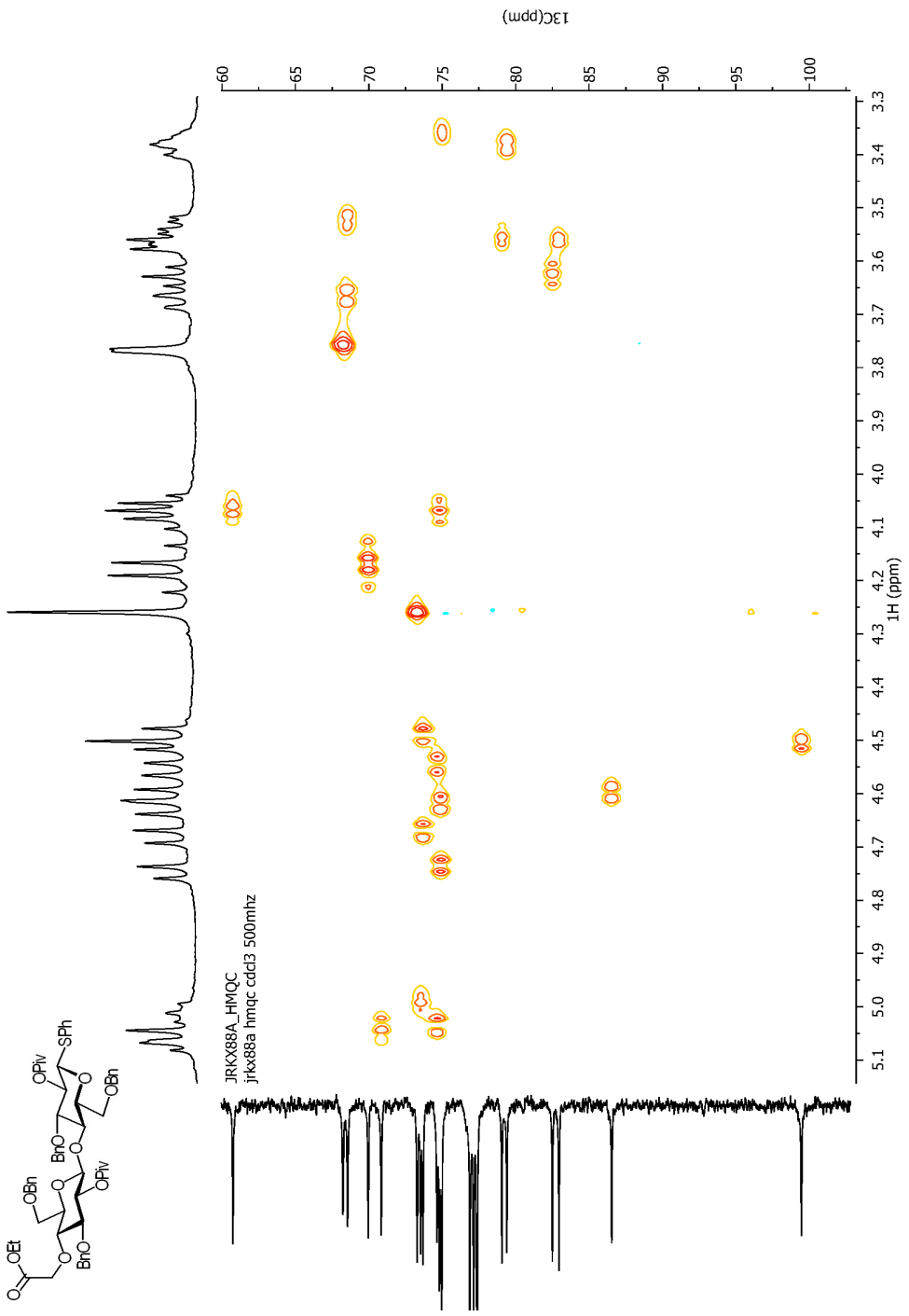


Figure 118: Compound 3.15 HMQC

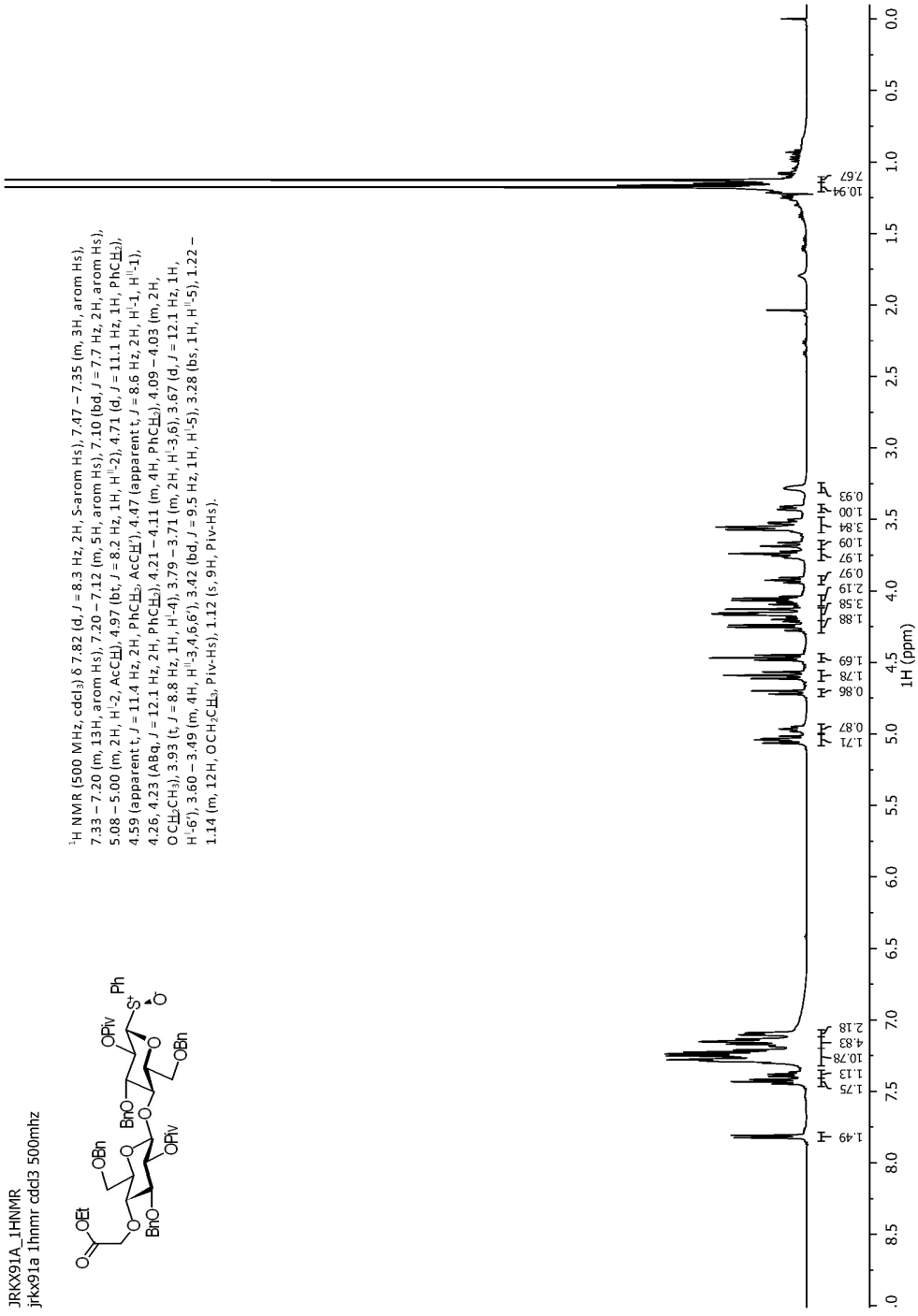


Figure 119: Compound 3.16a 1H-NMR

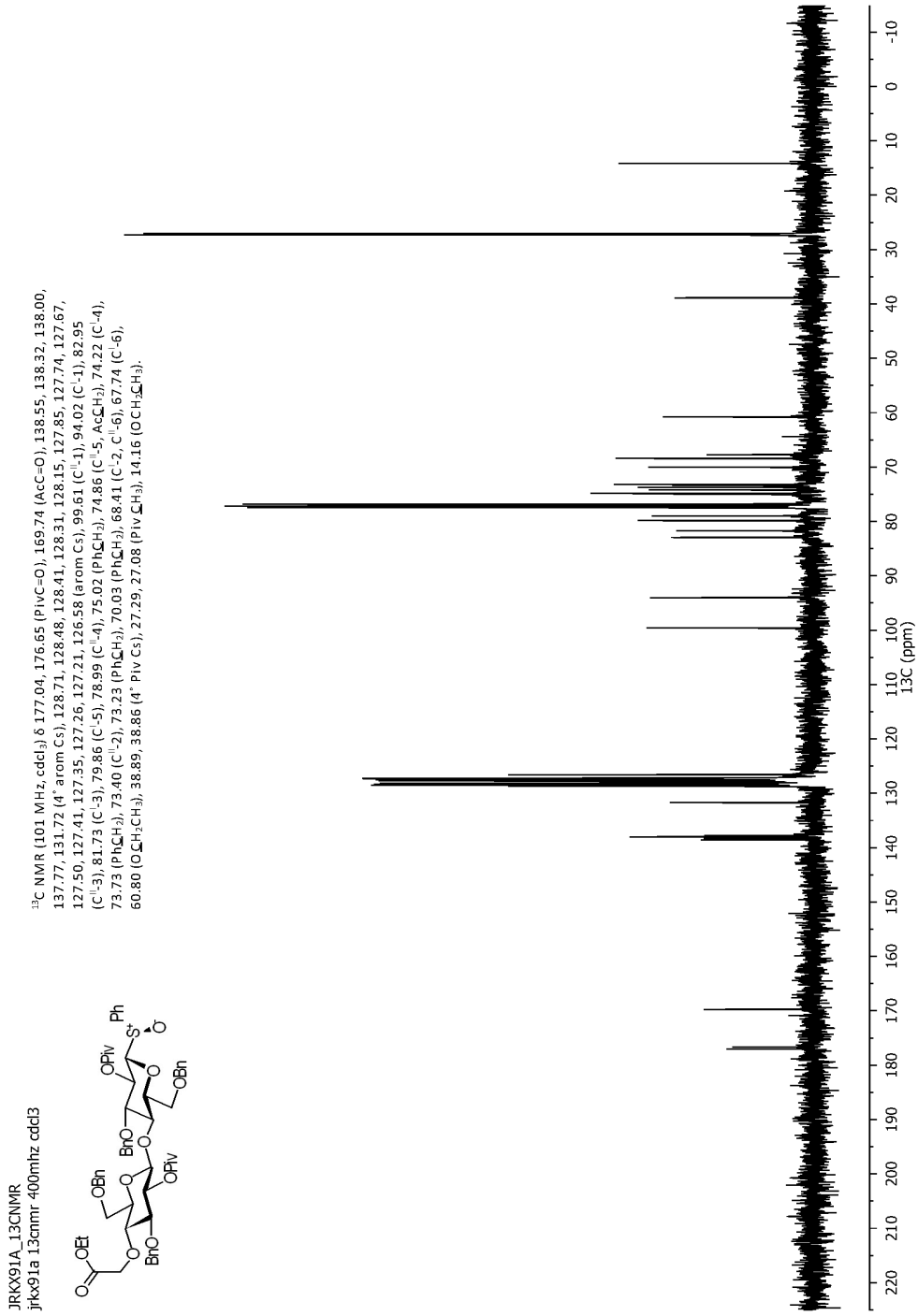


Figure 120: Compound 3.16a 13C-NMR

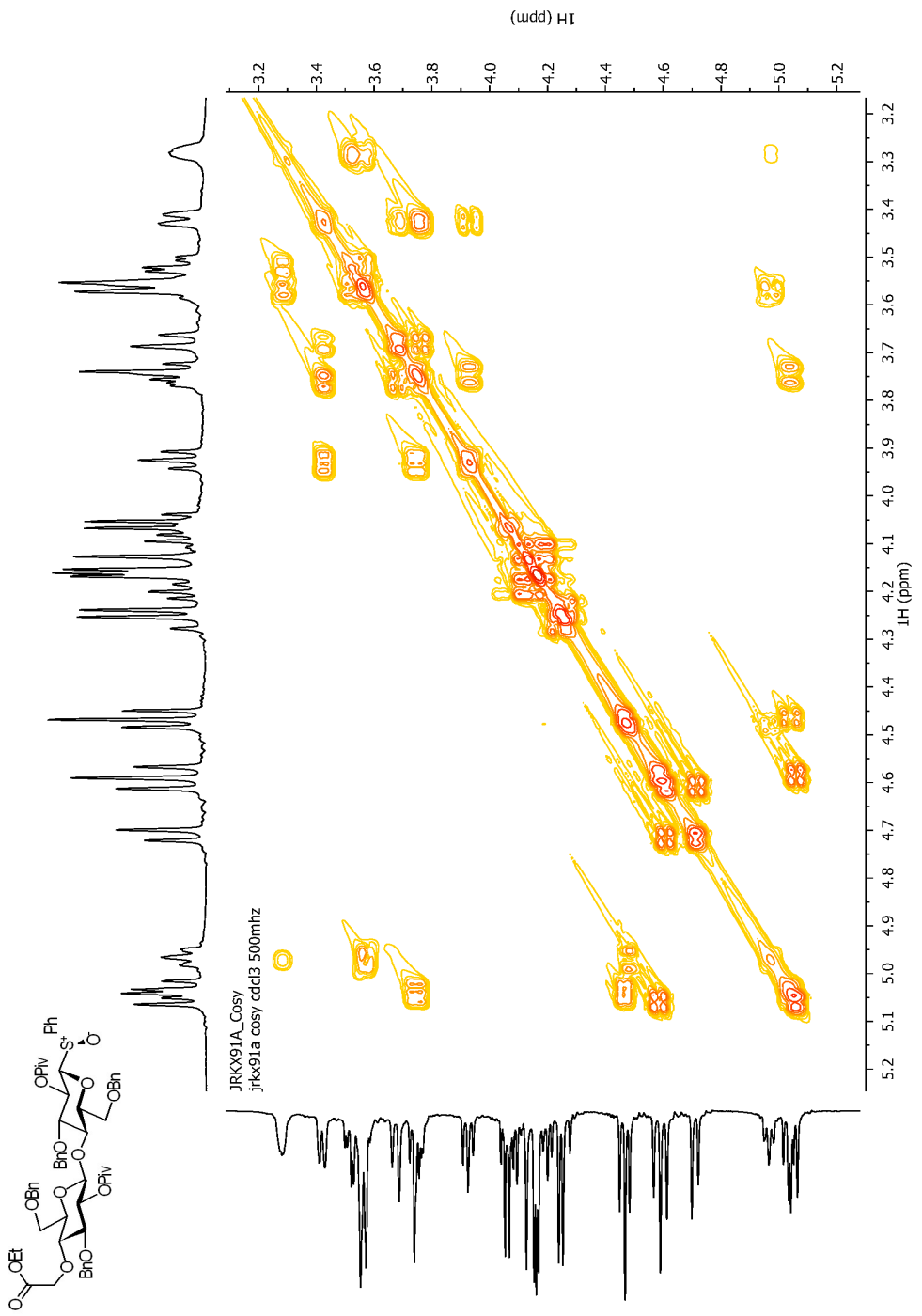


Figure 121: Compound 3.16a COSY

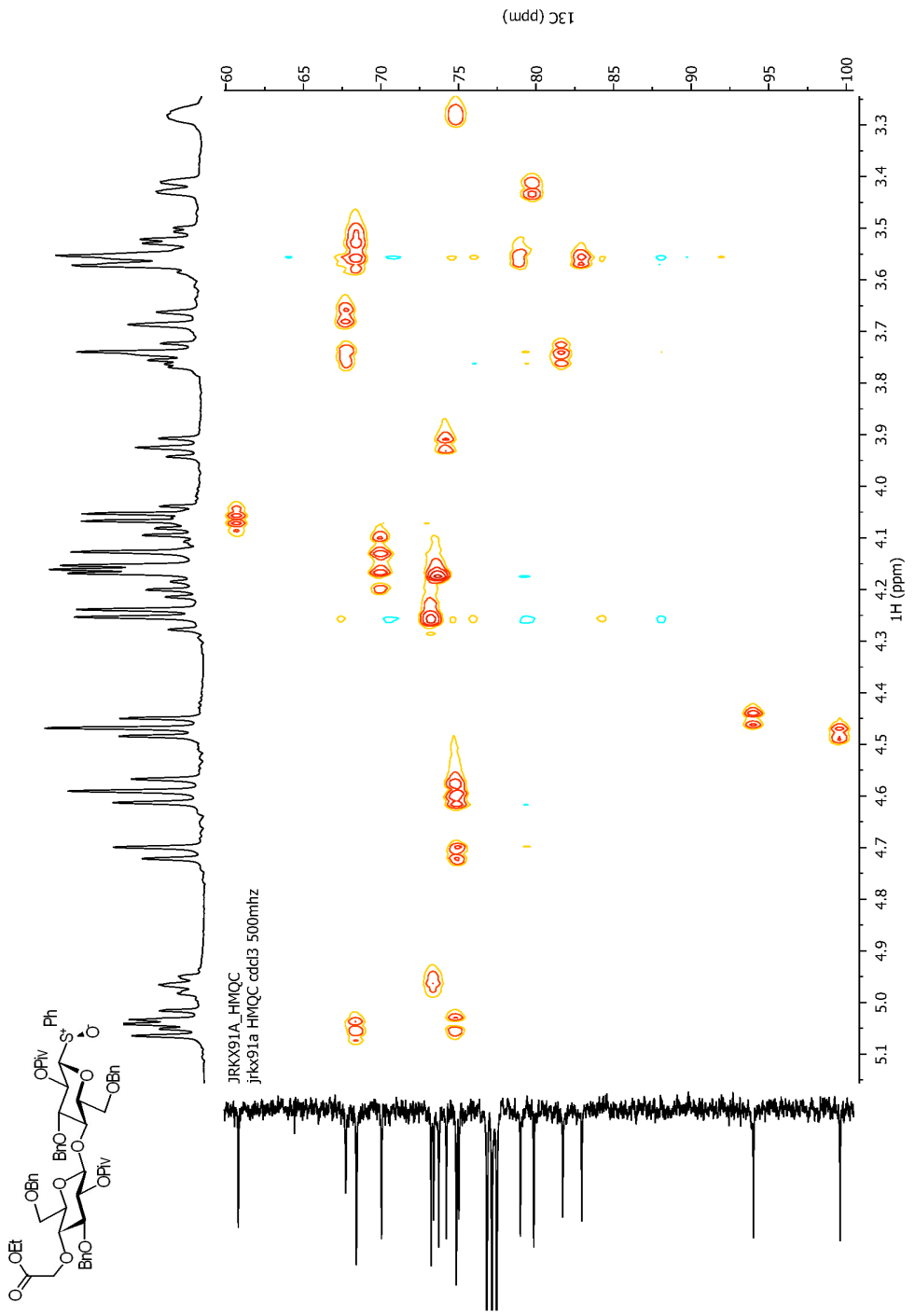


Figure 122: Compound 3.16a HMQC

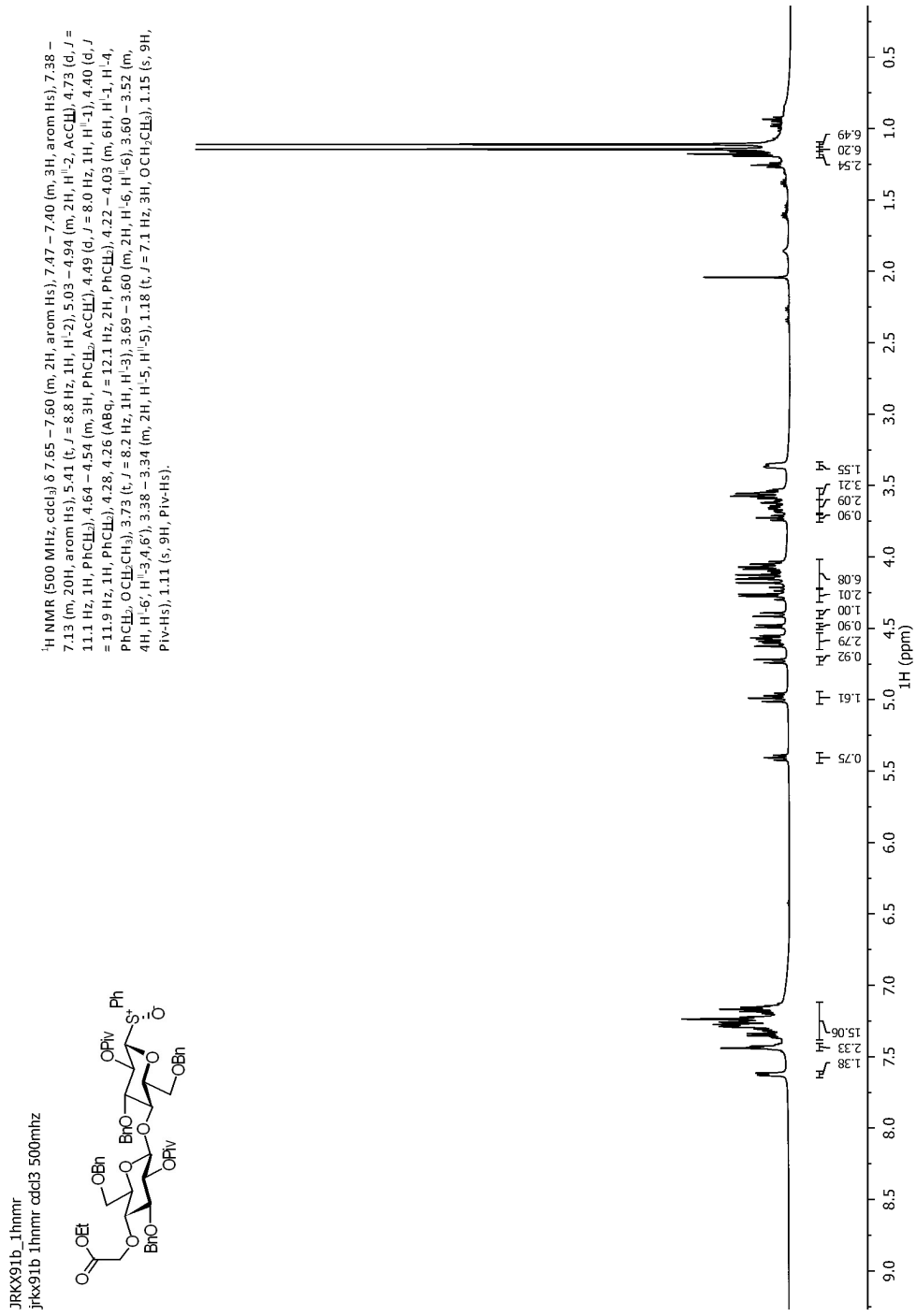


Figure 123: Compound 3.16b 1H-NMR

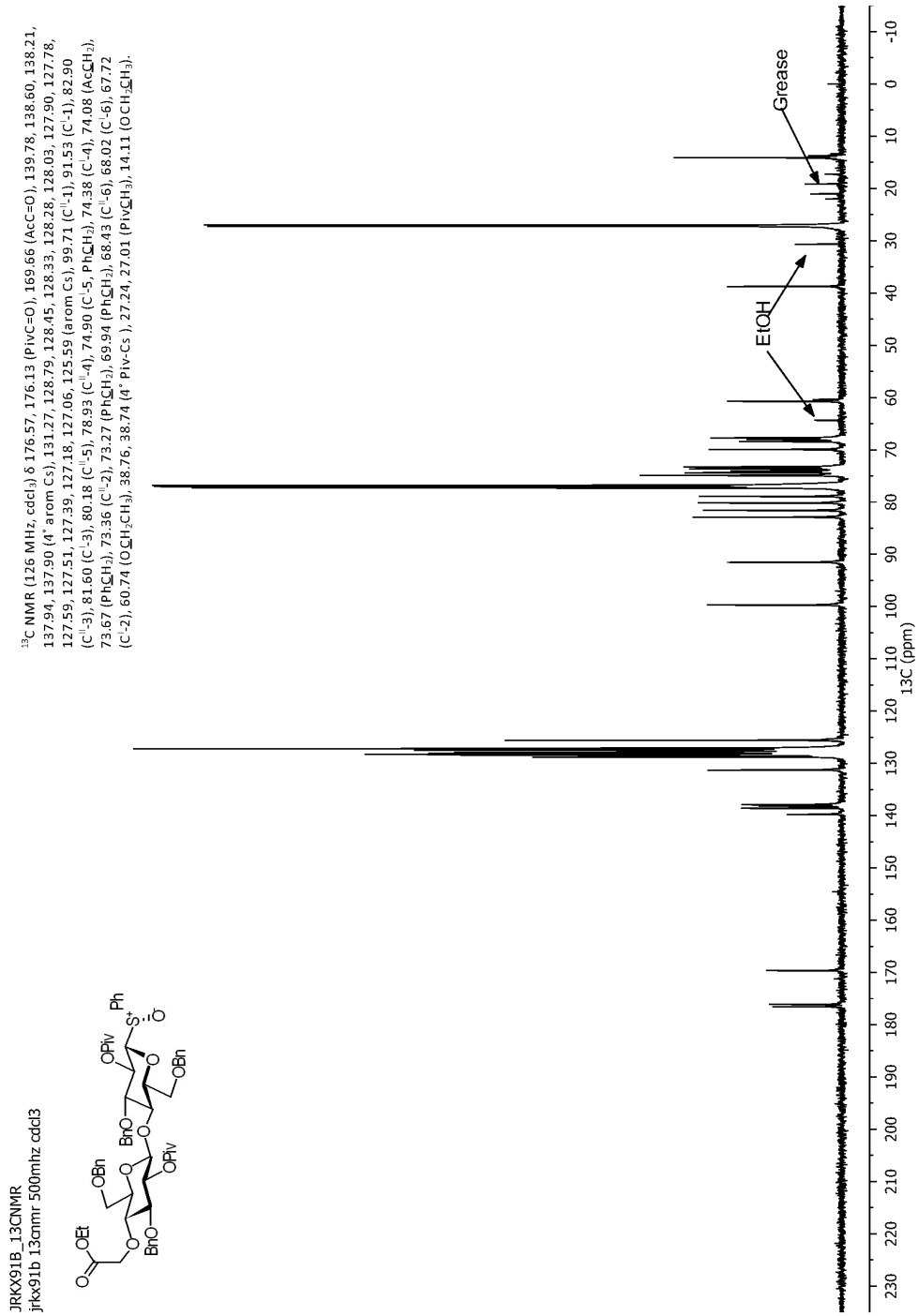


Figure 124: Compound 3.16b 13C-NMR

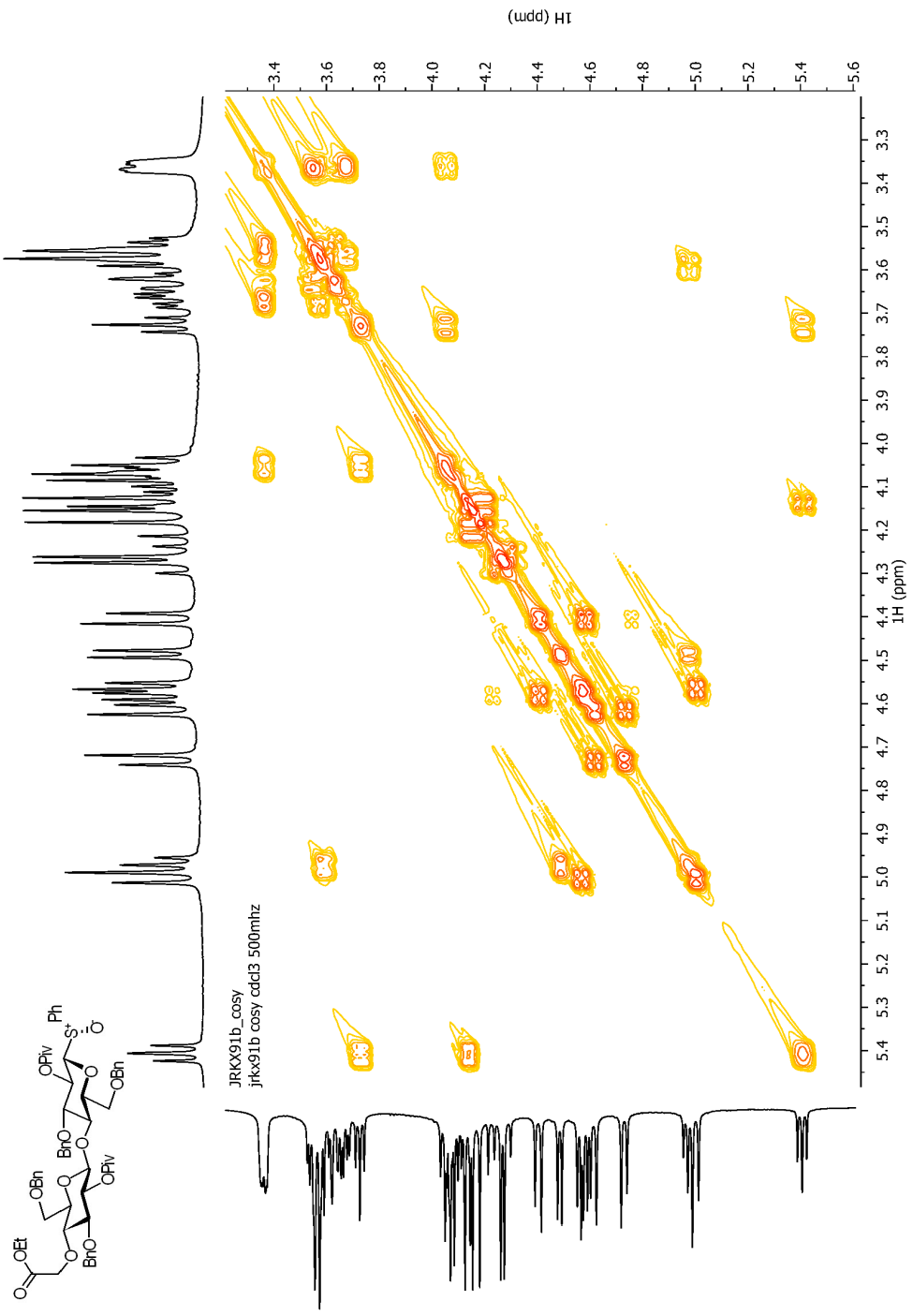


Figure 125: Compound 3.16b COSY

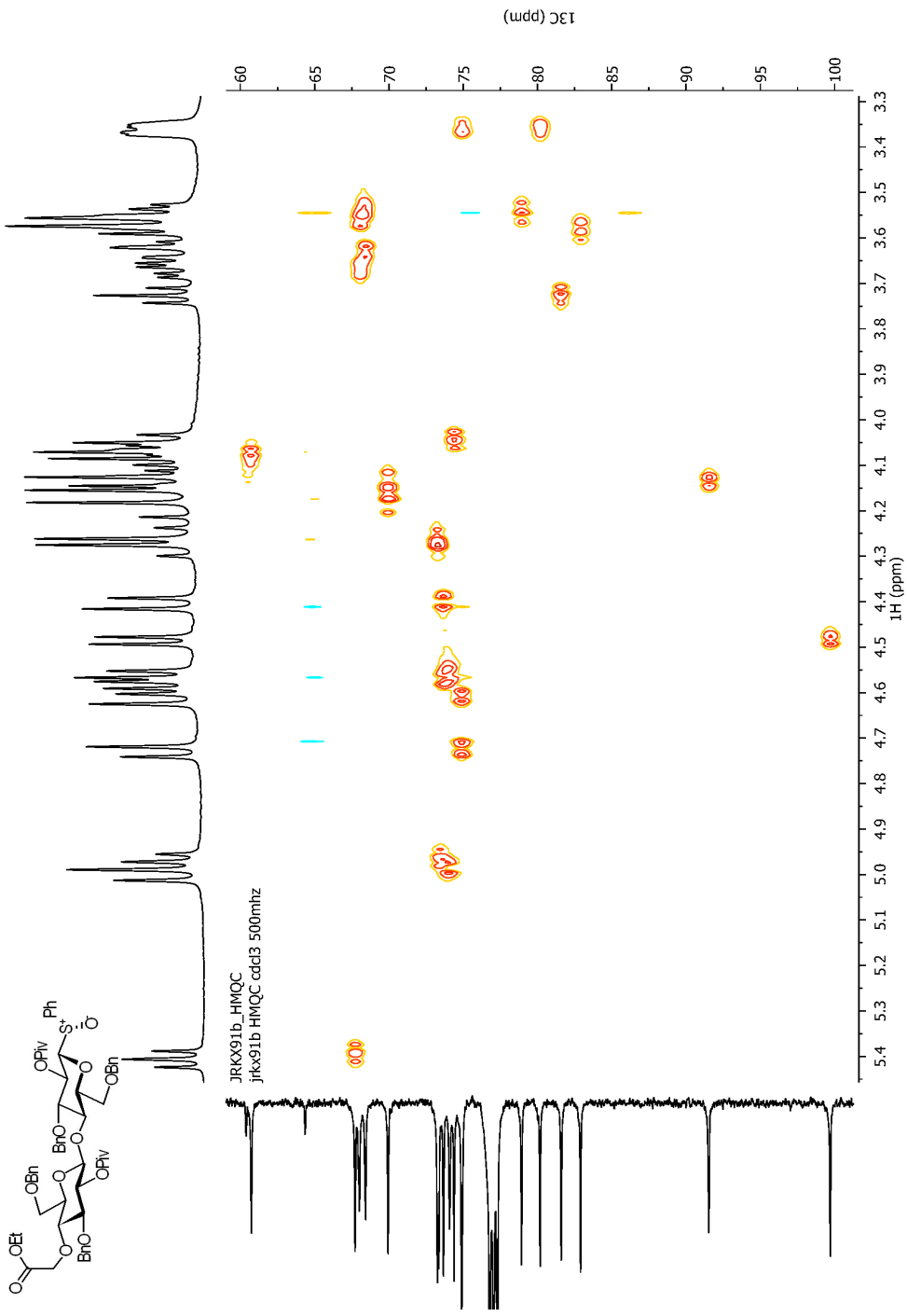


Figure 126: Compound 3.16b HMQC

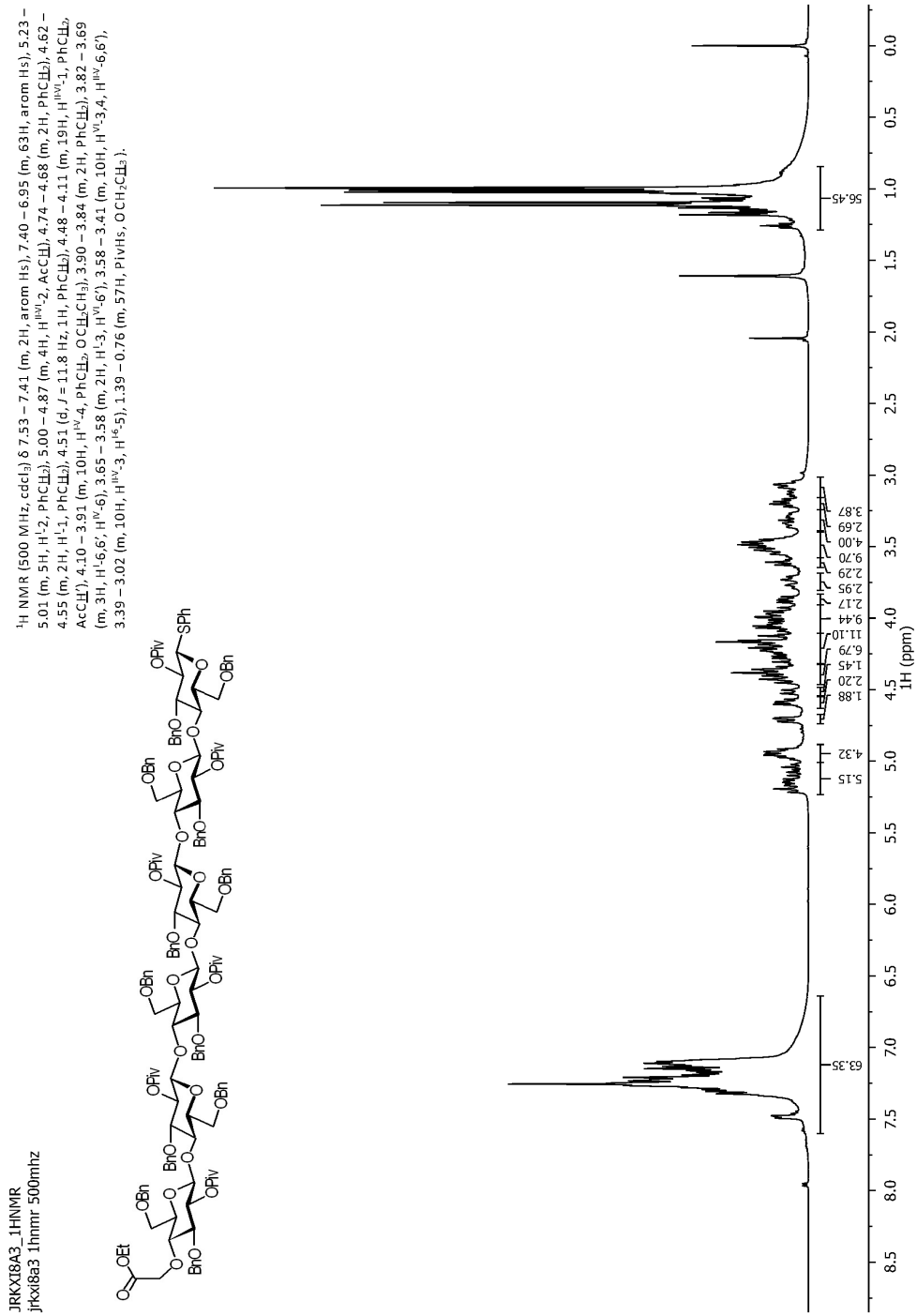


Figure 127: Compound 3.17 1H-NMR

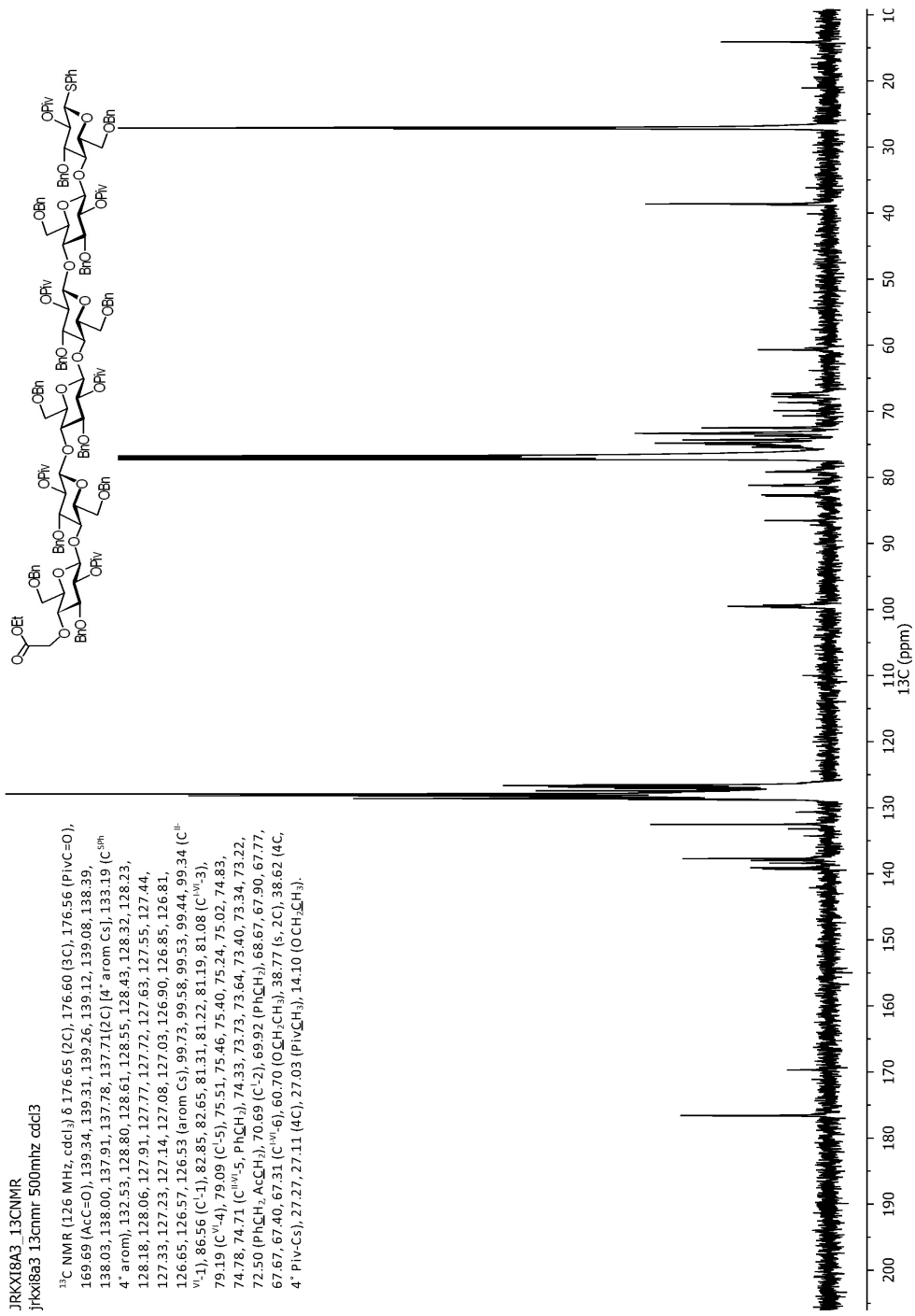


Figure 128: Compound 3.17 13C-NMR

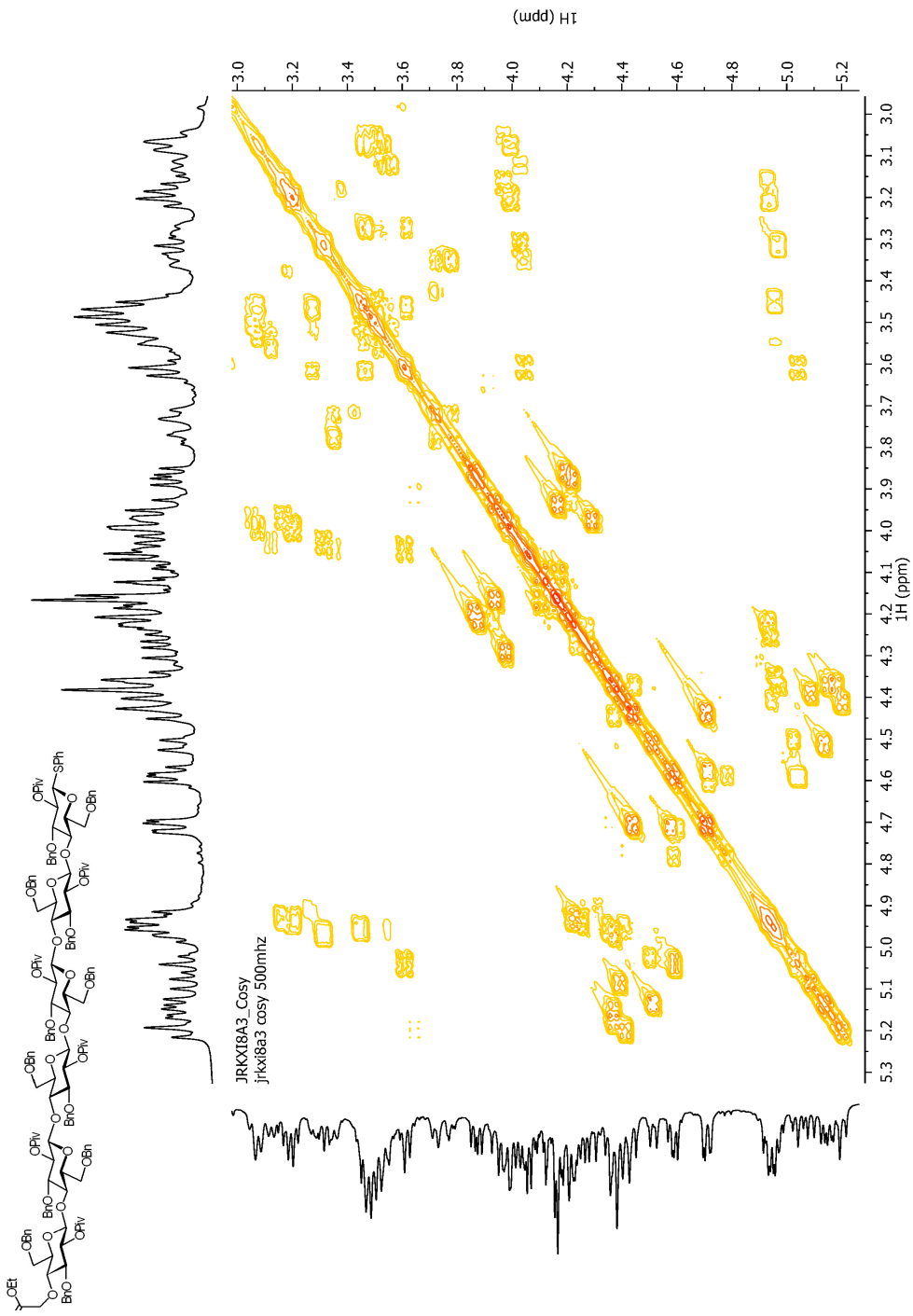


Figure 129: Compound 3.17 COSY

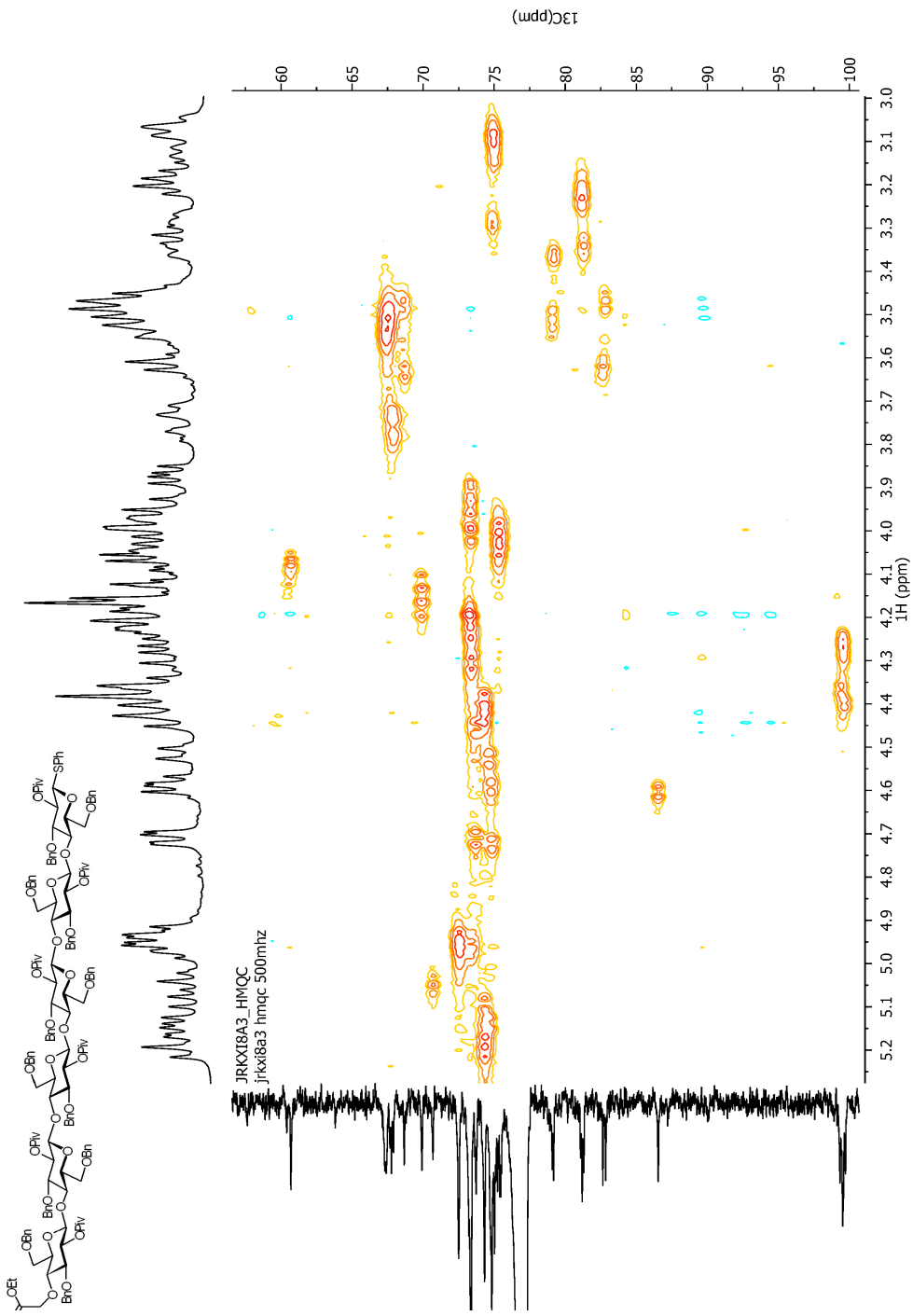


Figure 130: Compound 3.17 HMQC

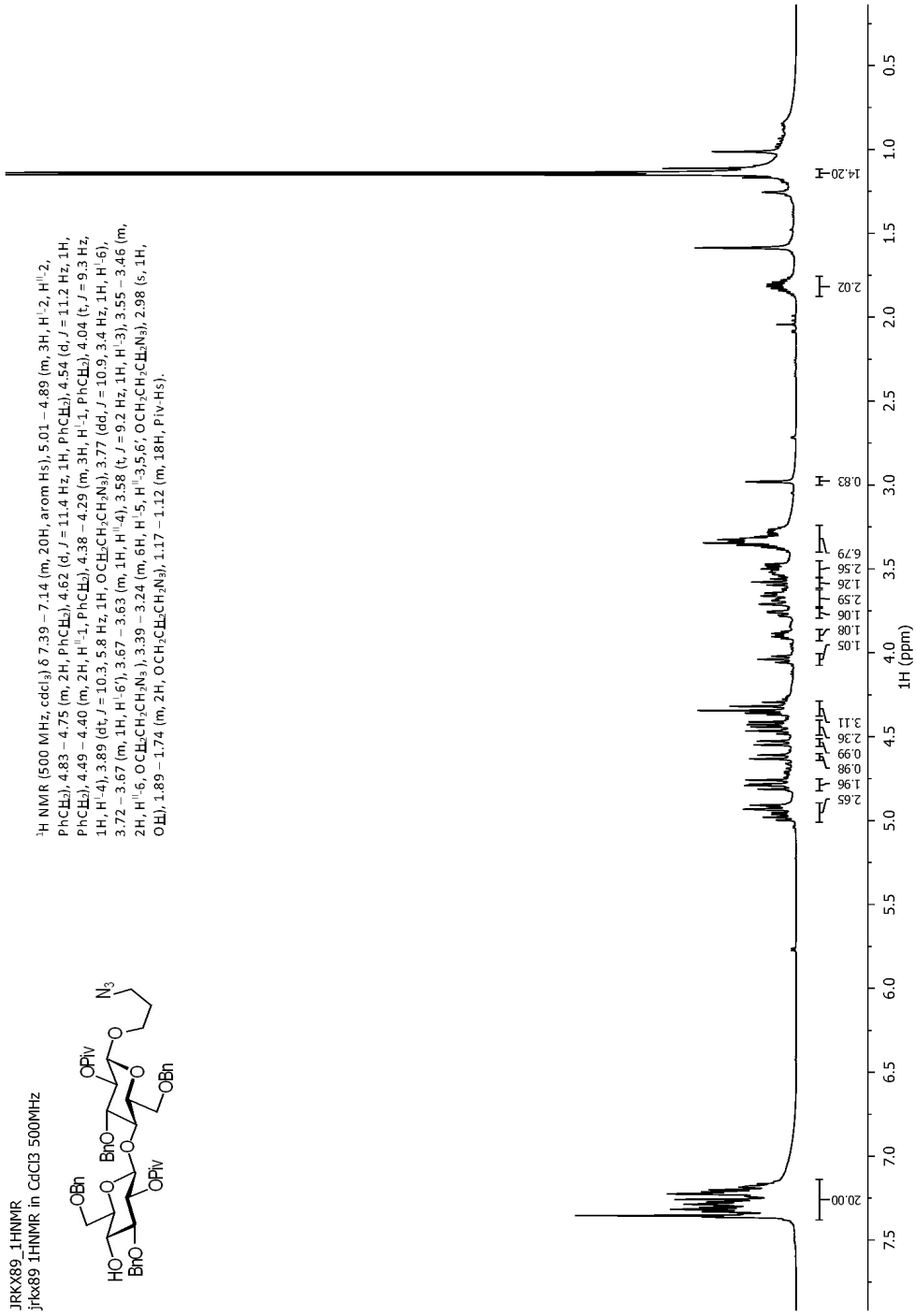


Figure 131: Compound 3.18 1H-NMR

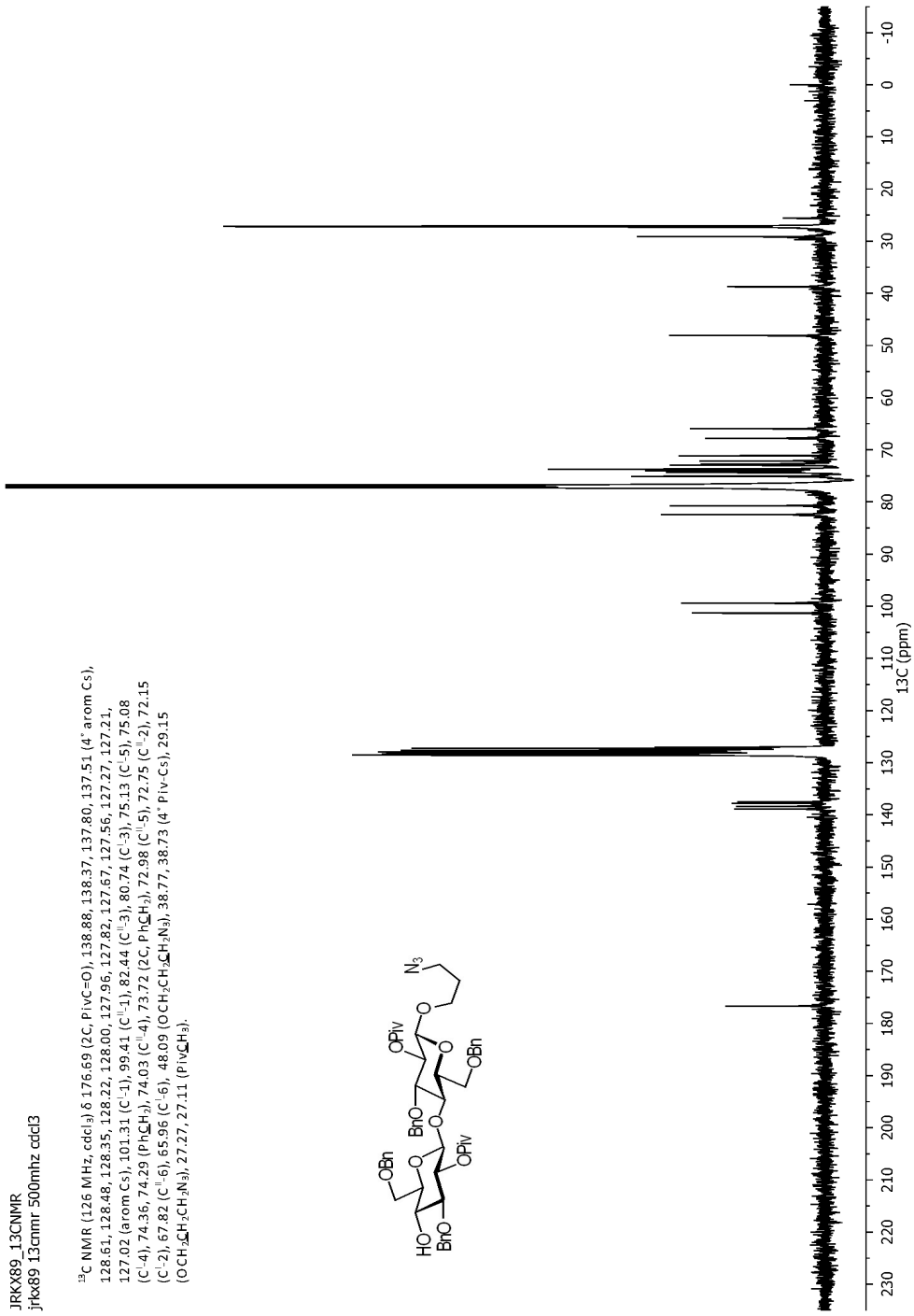


Figure 132: Compound 3.18 13C-NMR

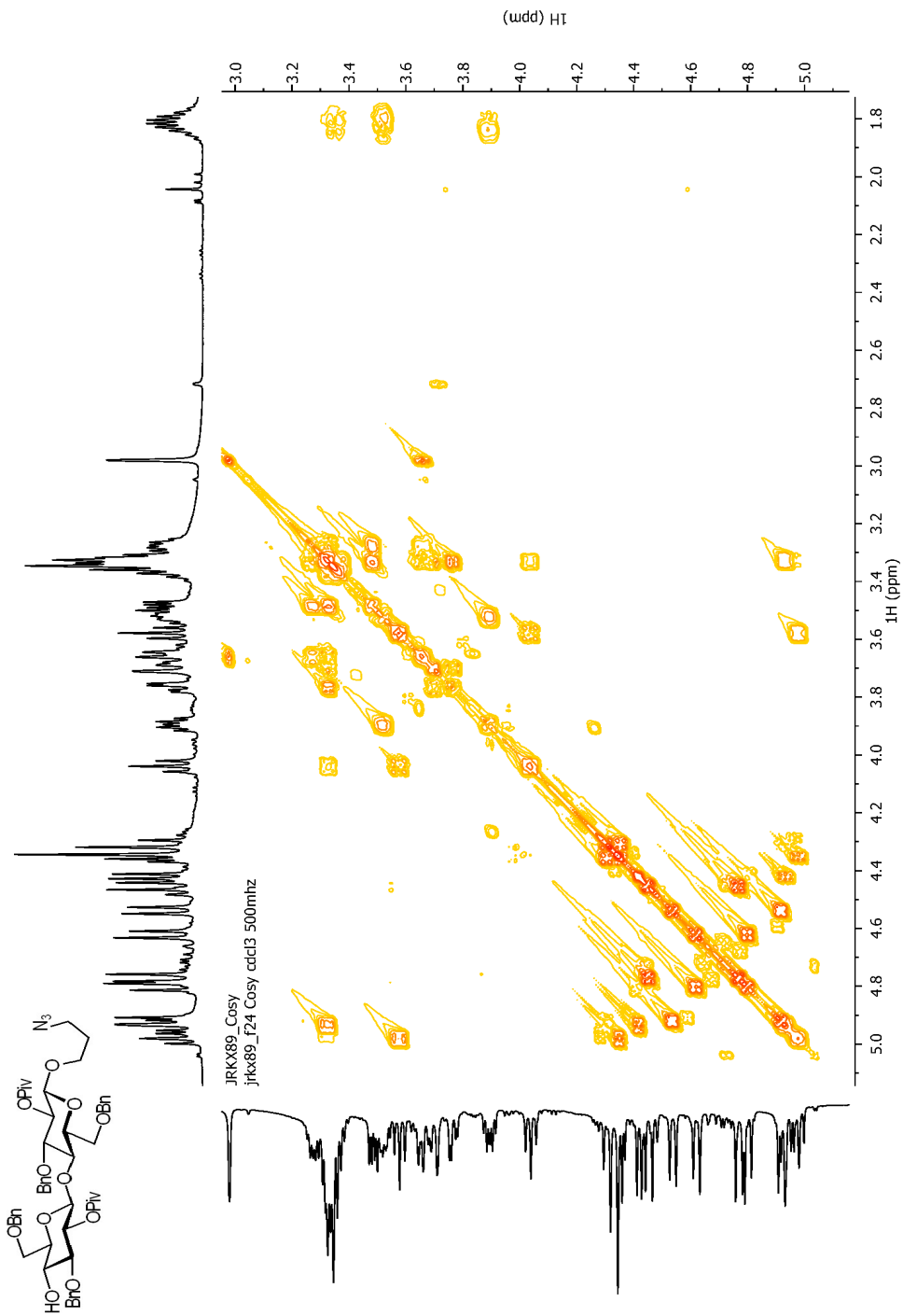


Figure 133: Compound 3.18 COSY

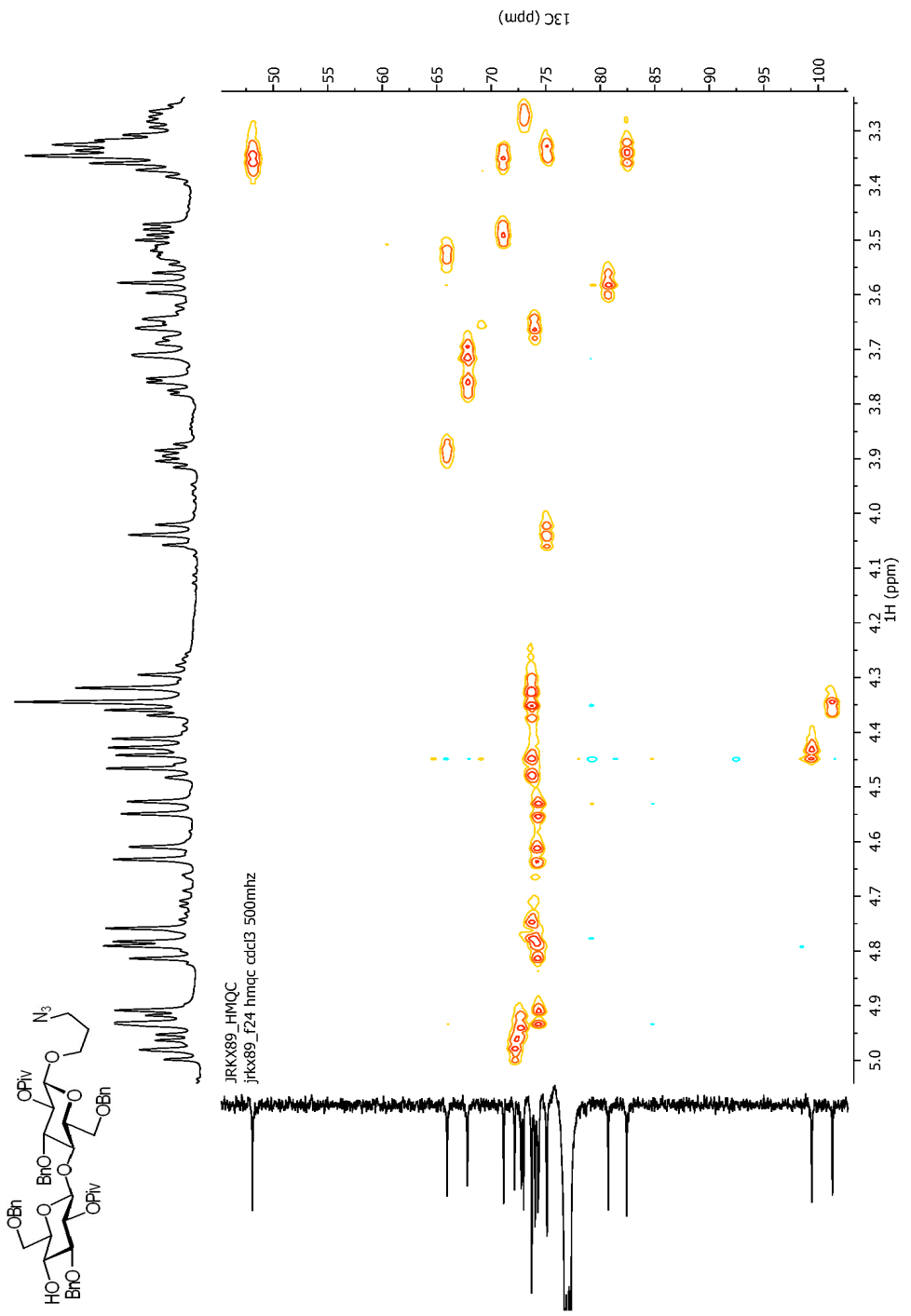


Figure 134: Compound 3.18 HMQC

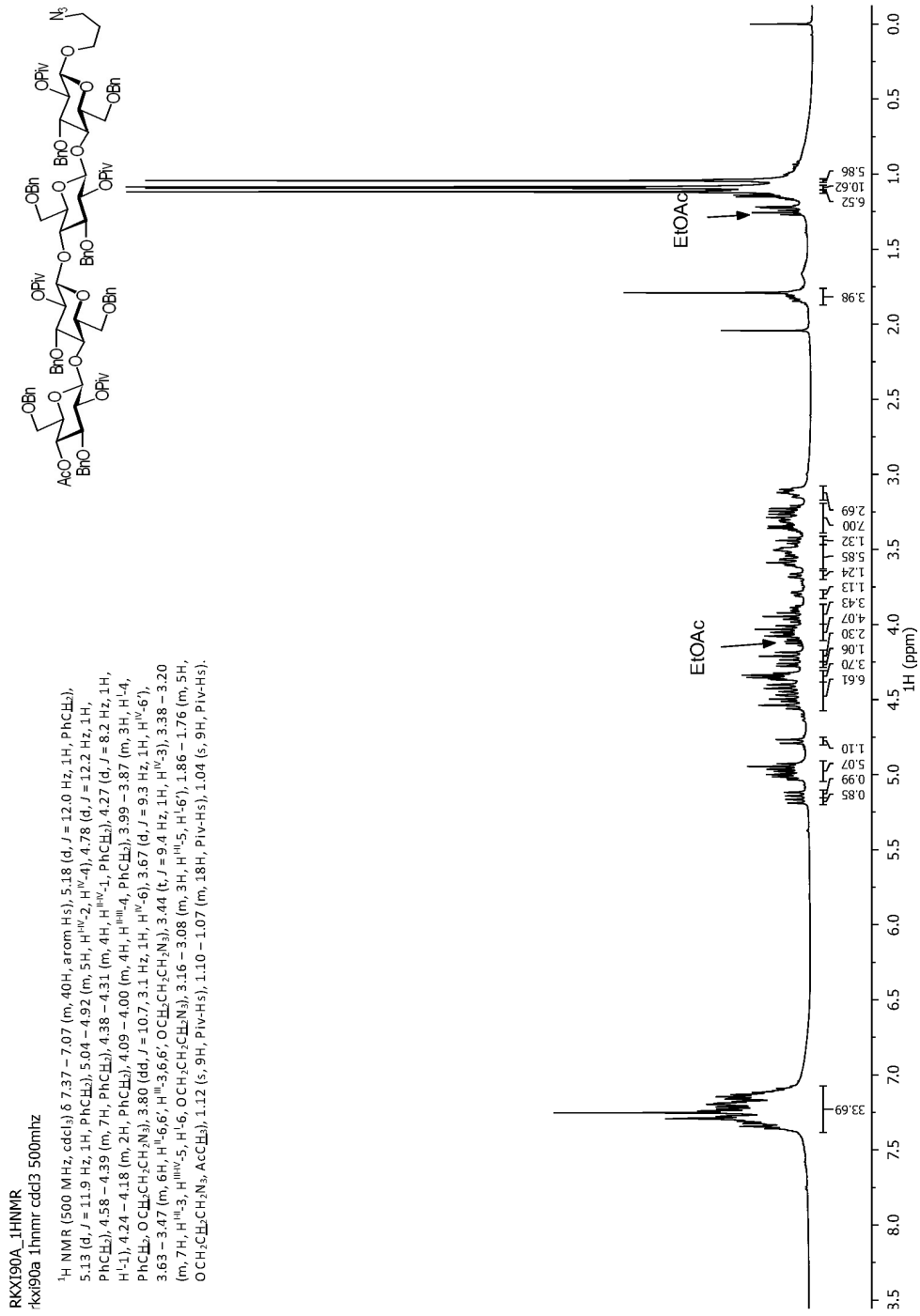


Figure 135: Compound 3.19 1H-NMR

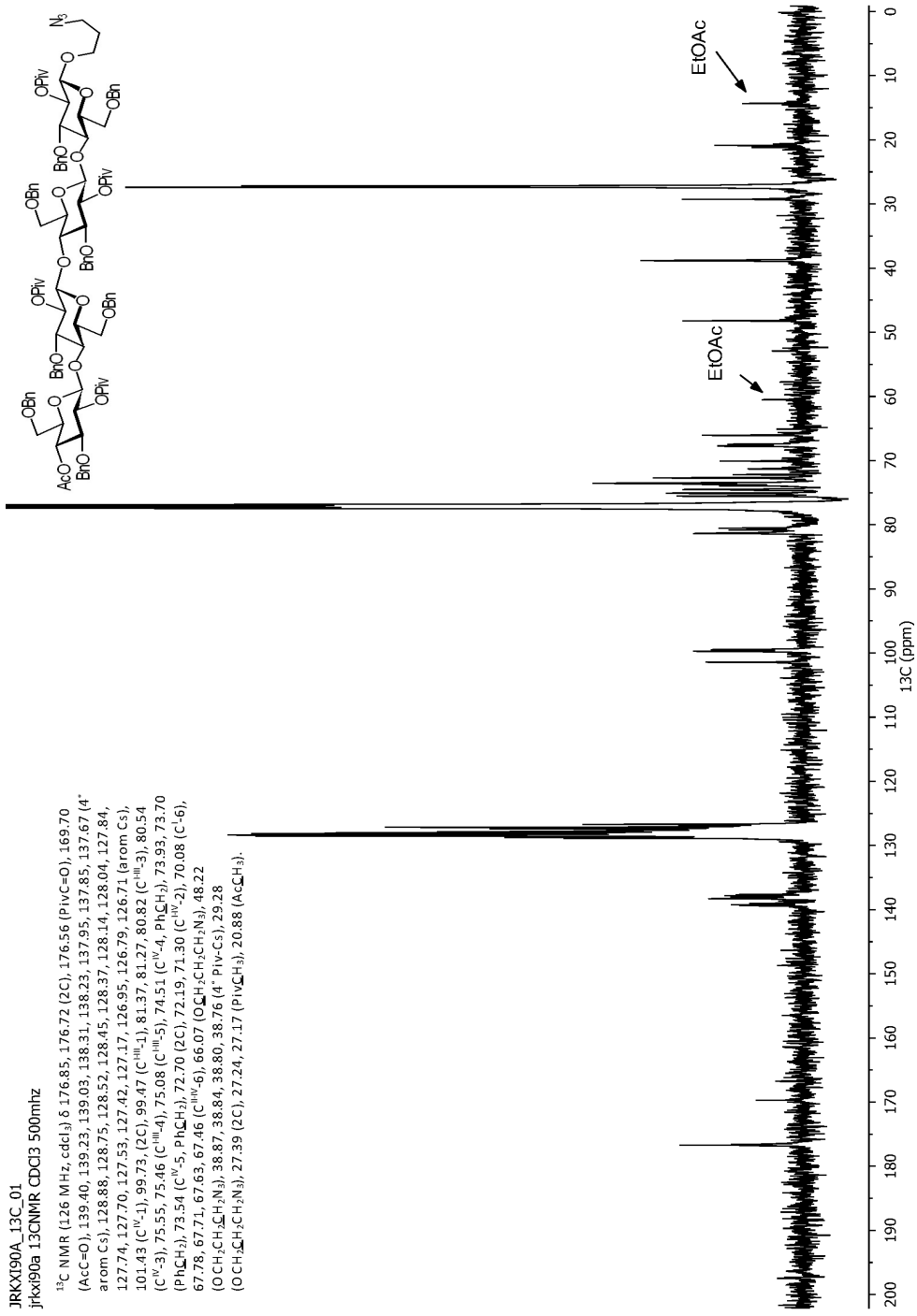


Figure 136: Compuond 3.19 13C-NMR

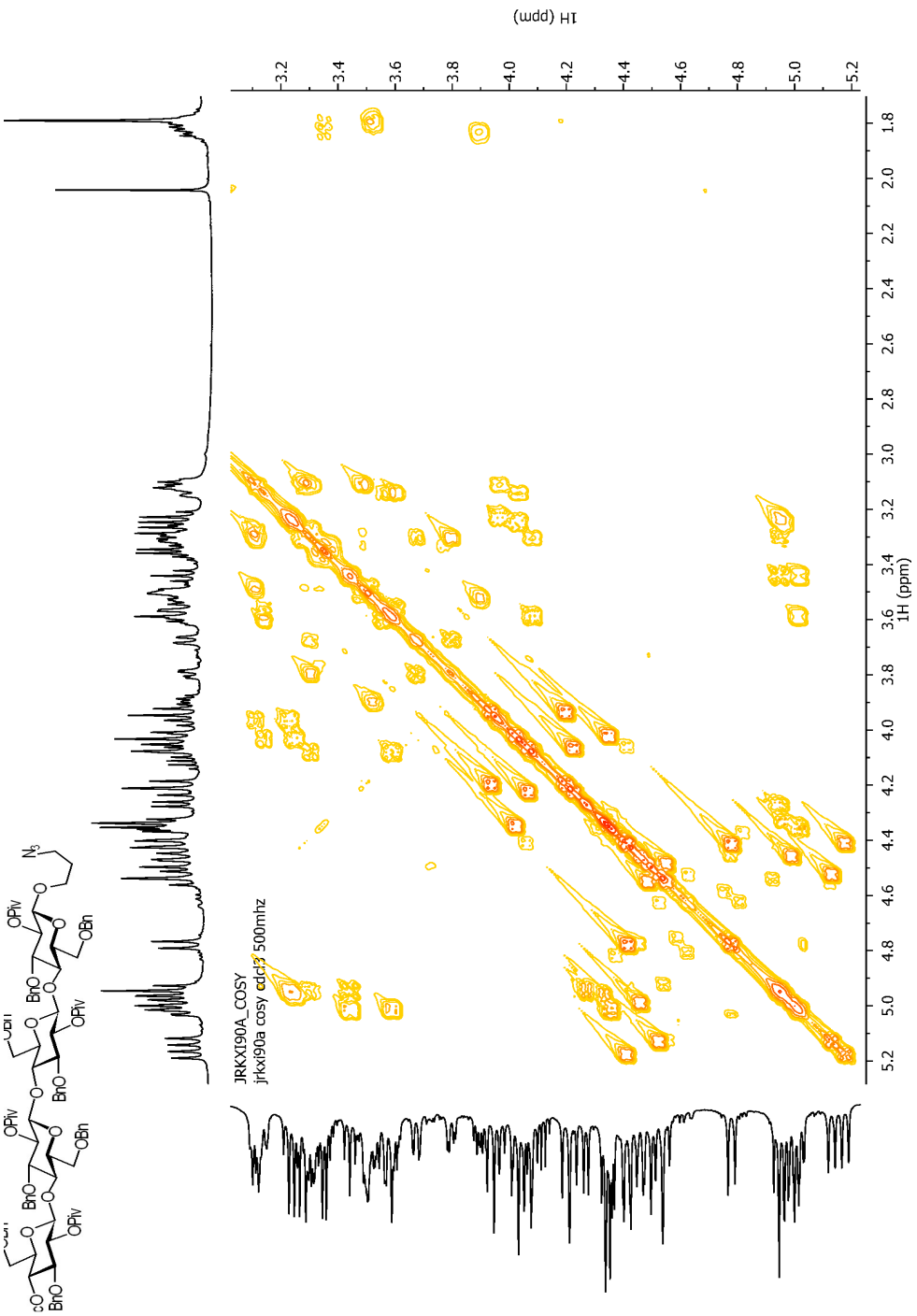


Figure 137: Compound 3.19 COSY

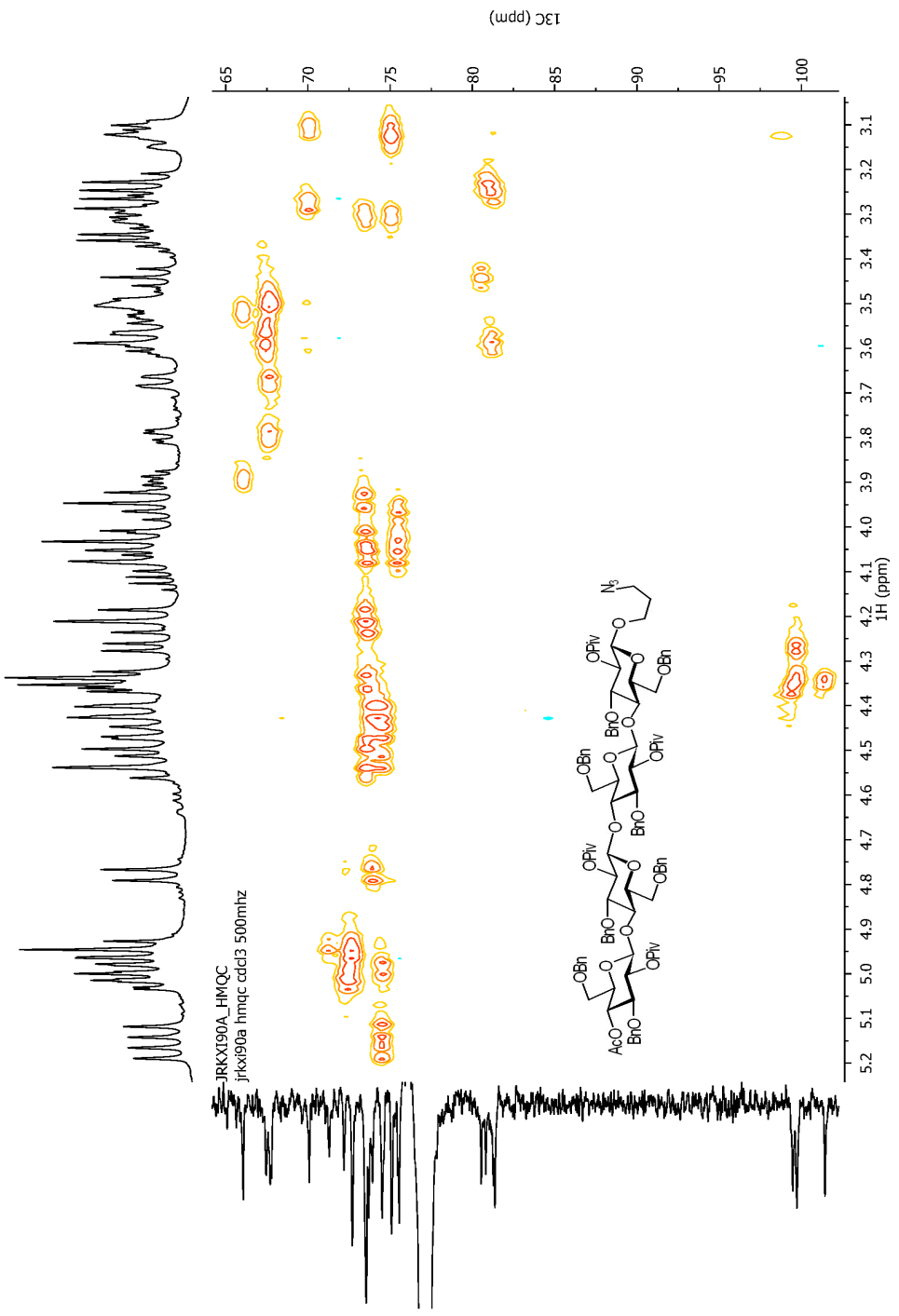


Figure 138: Compound 3.19 HMQC

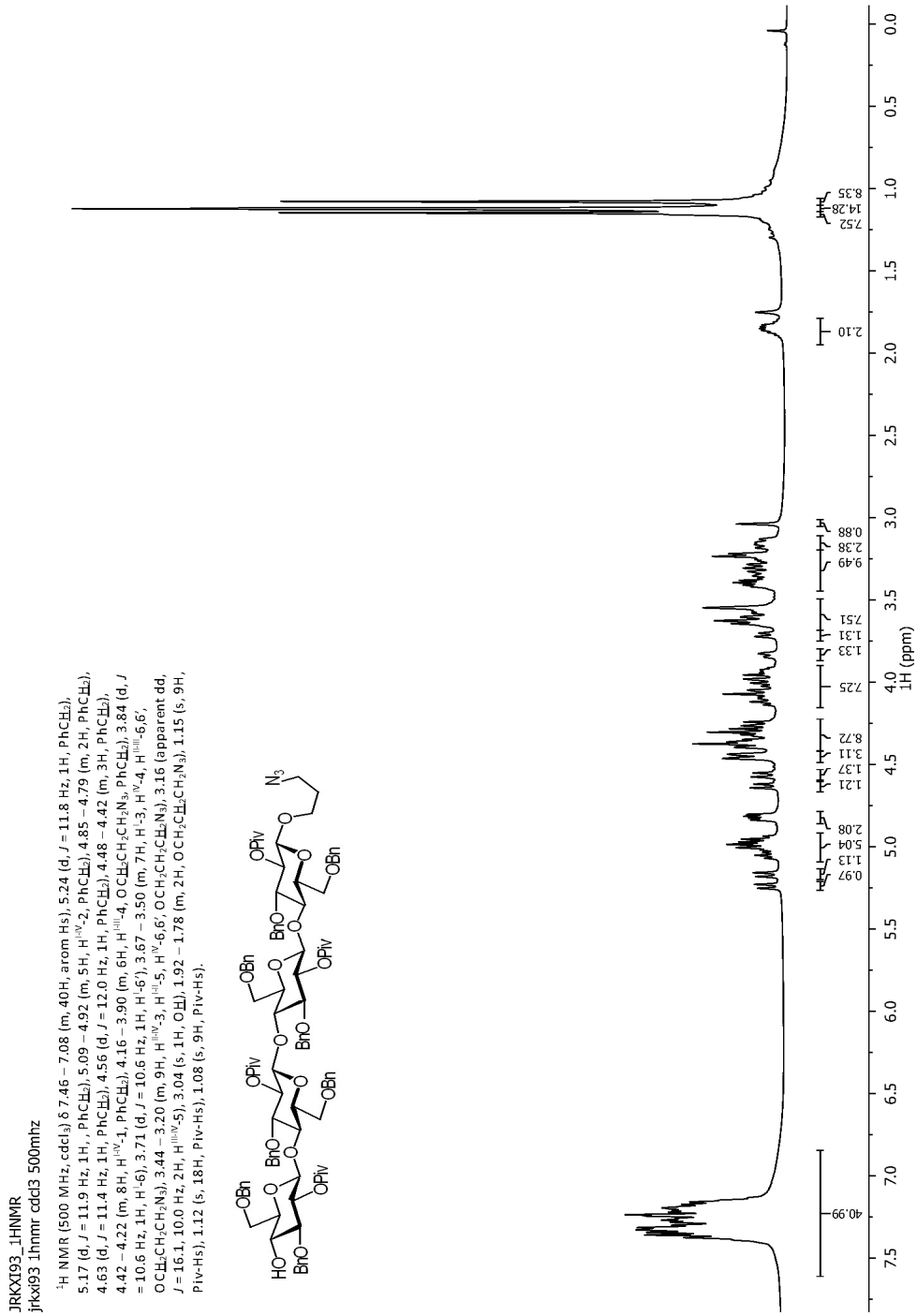


Figure 139: Compound 3.20 1H-NMR

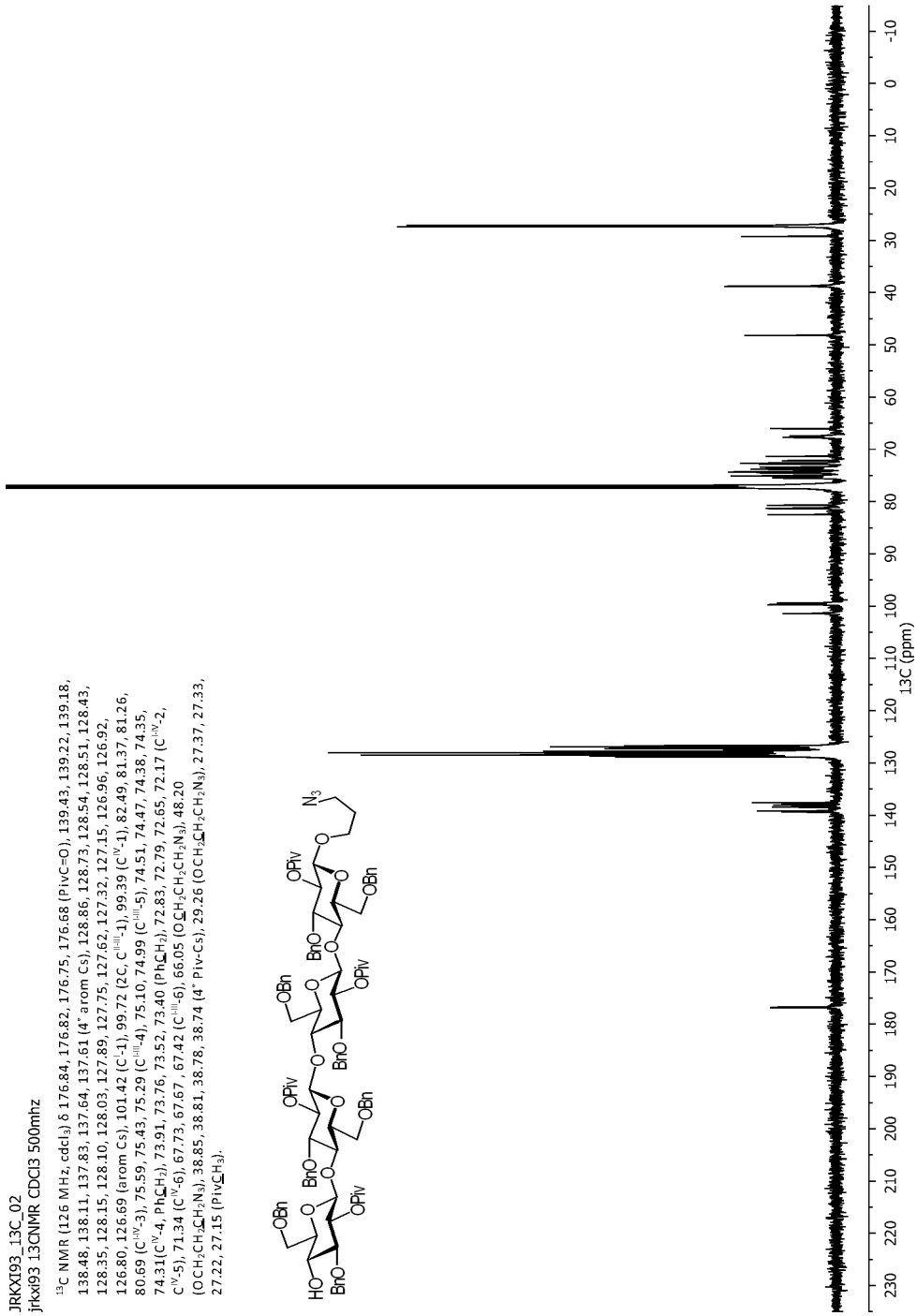


Figure 140: Compound 3.20 13C-NMR

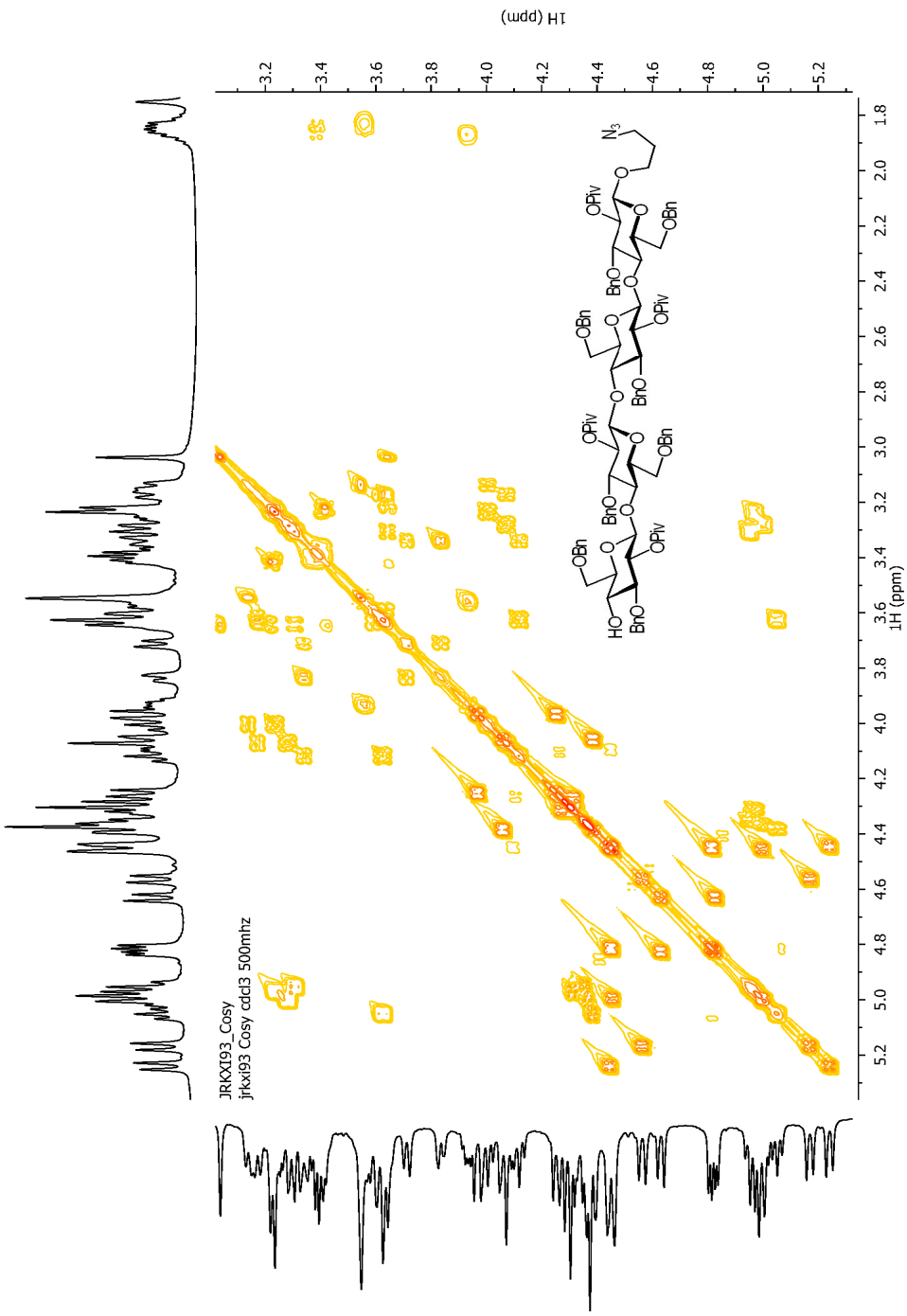


Figure 141: Compuond 3.20 COSY

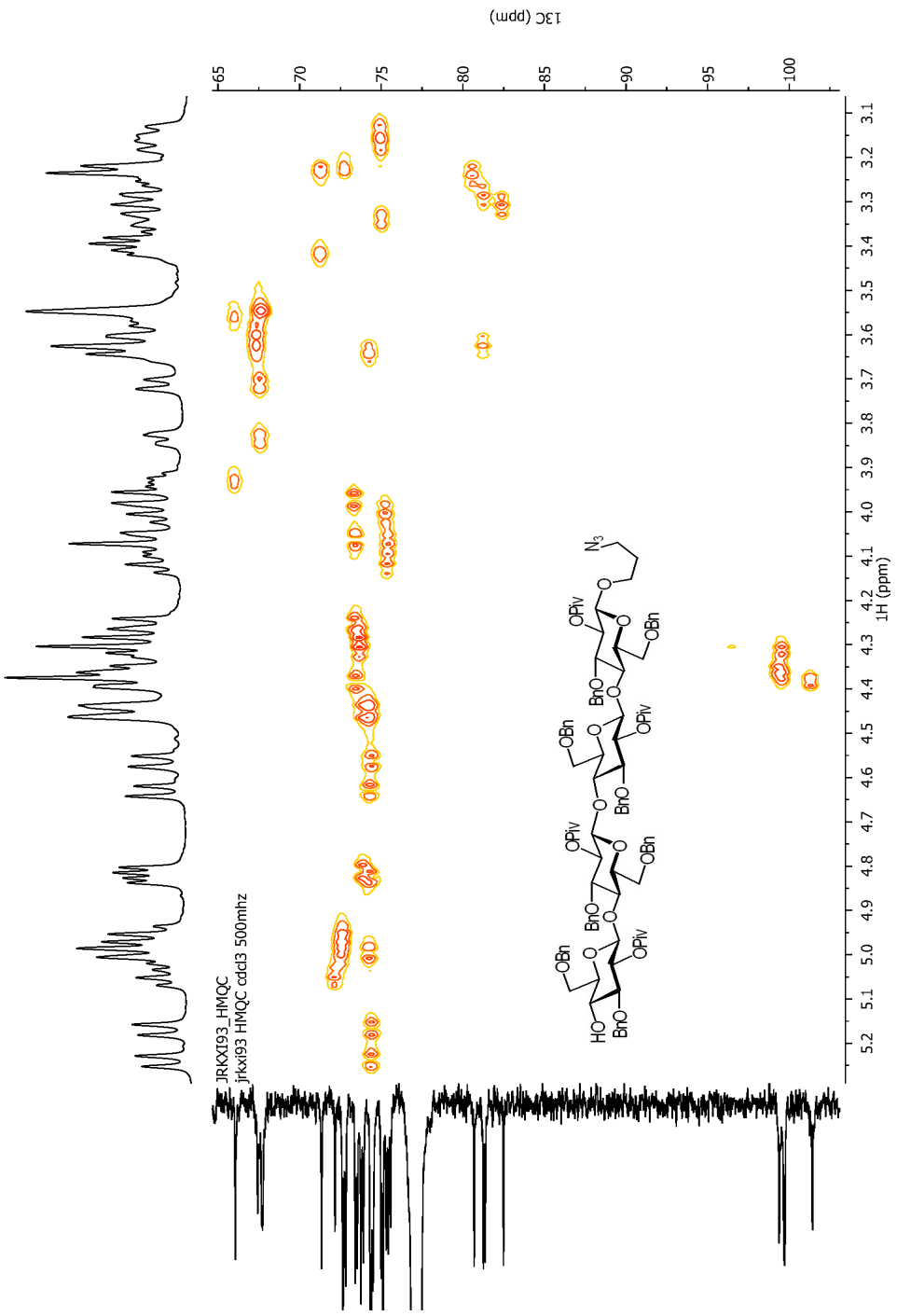


Figure 142: Compound 3.20 HMQC

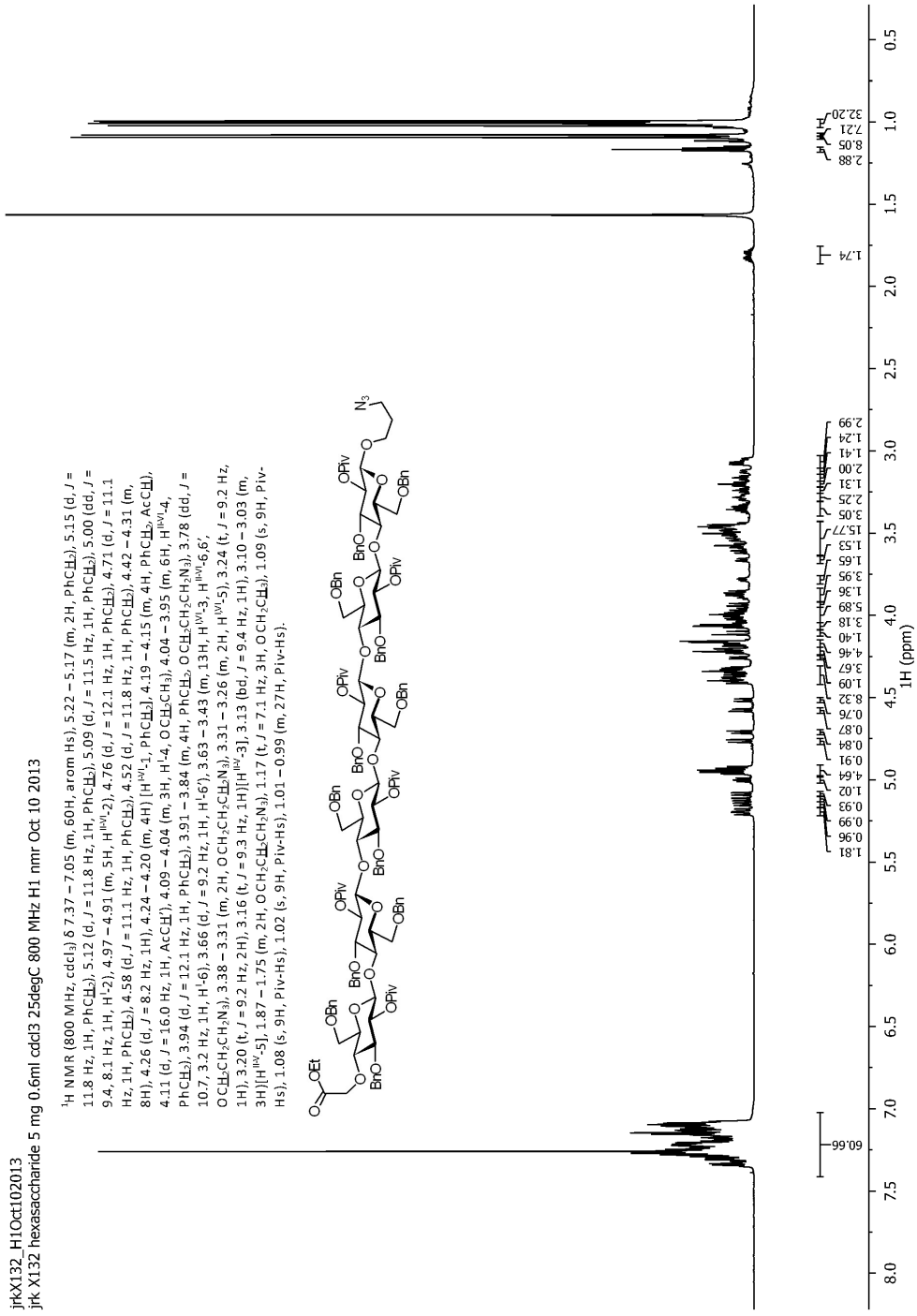


Figure 143: Compound 3.21 1H-NMR

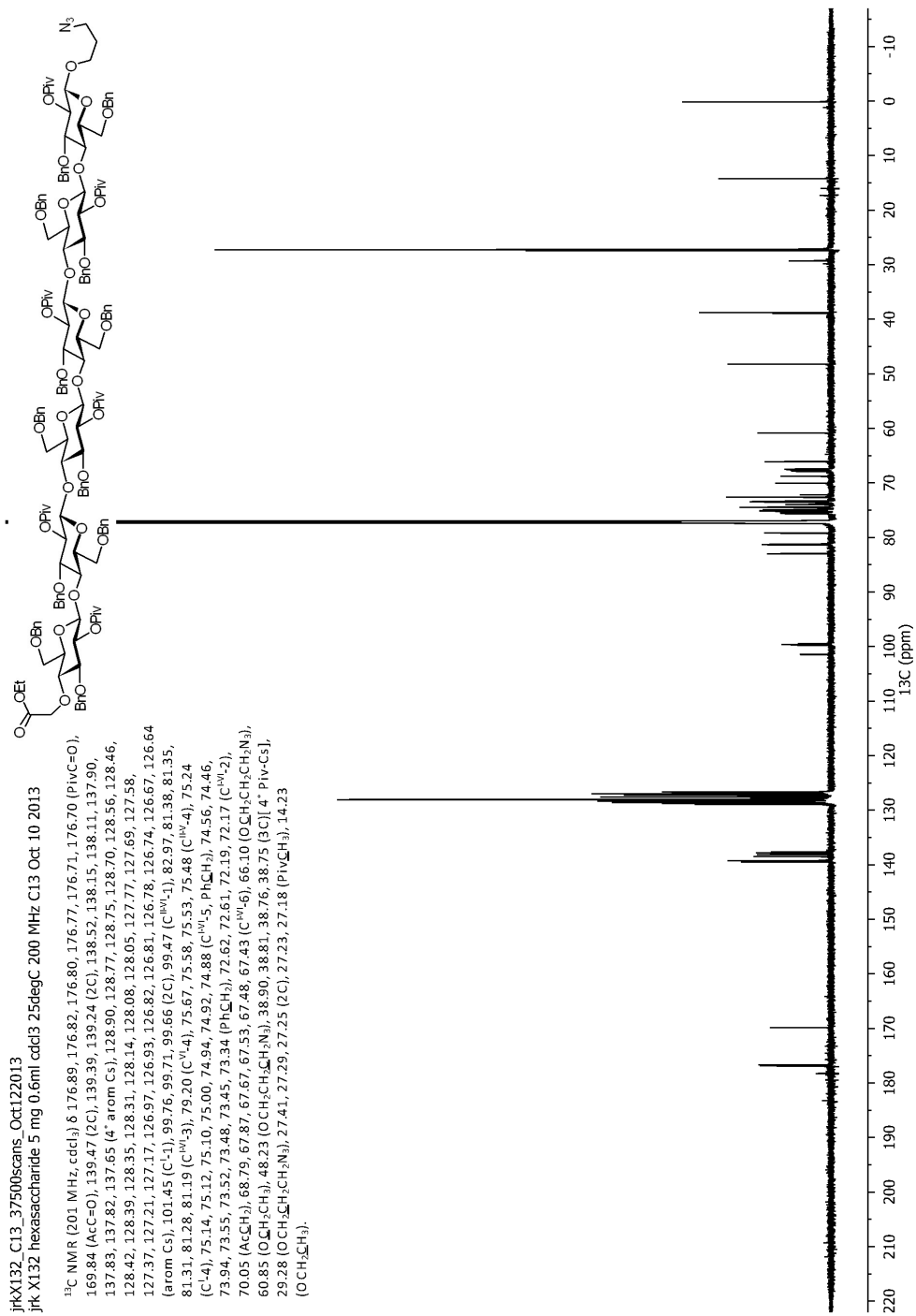


Figure 144: Compound 3.21 13C-NMR

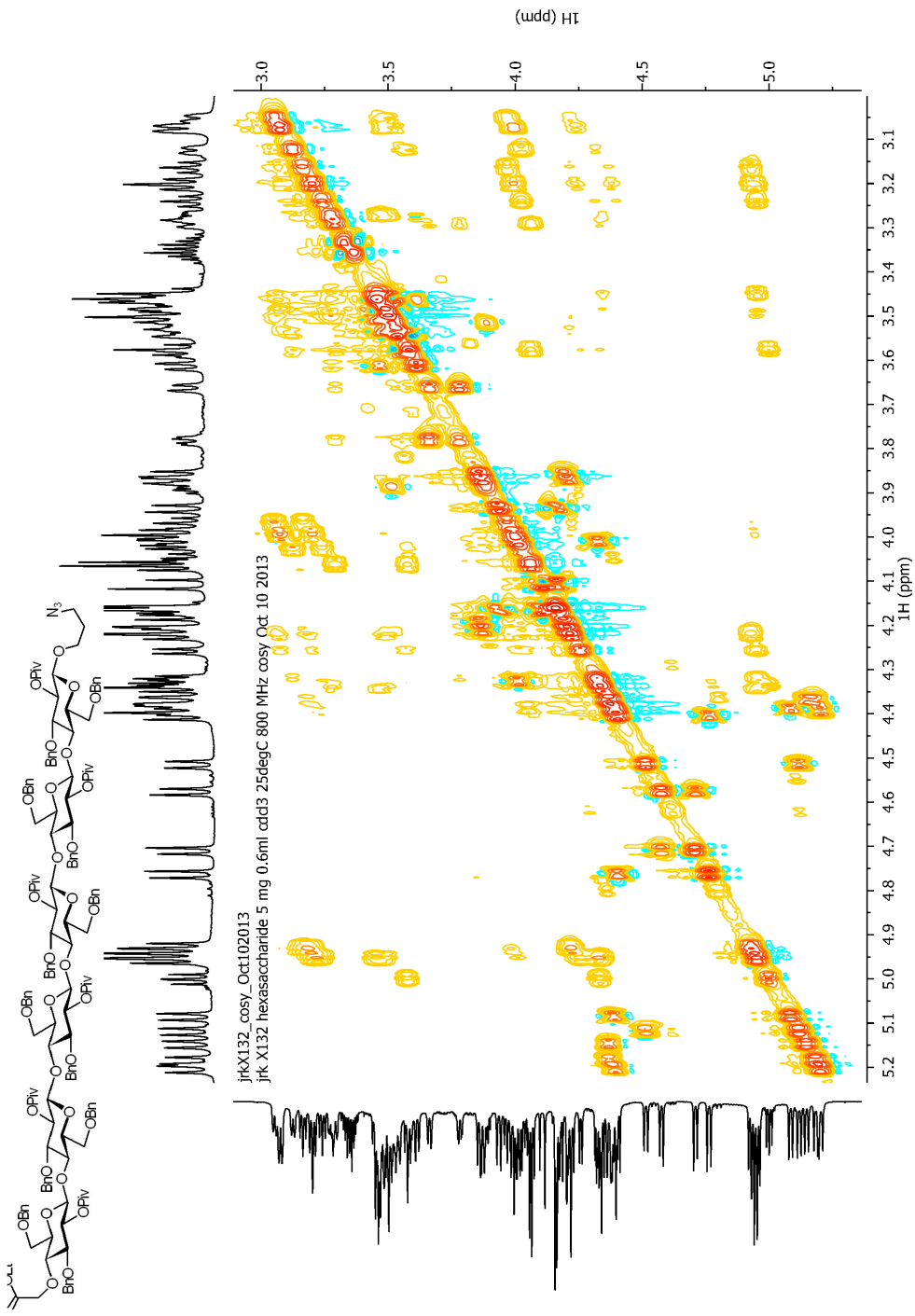


Figure 145: Compound 3.21 COSY

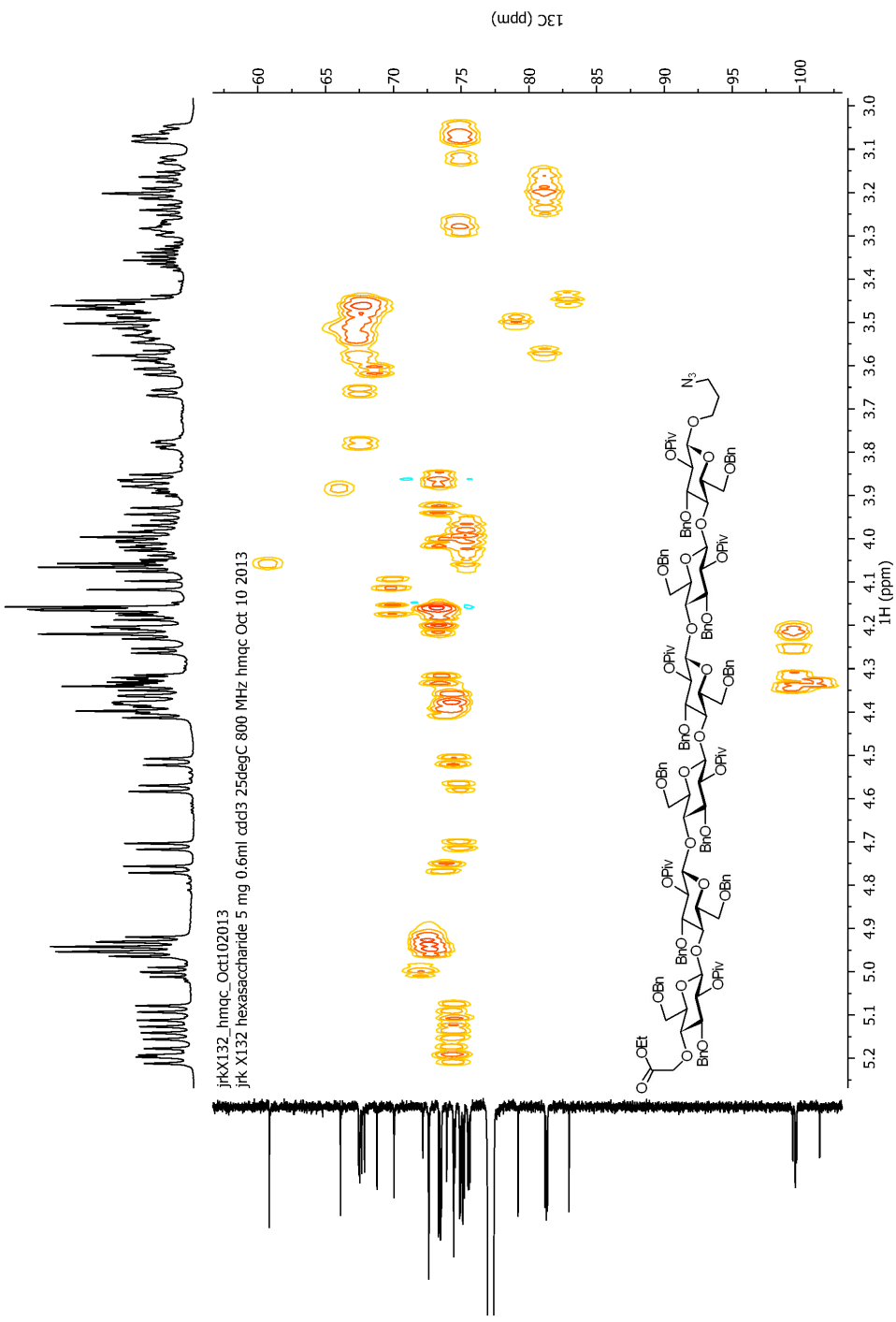


Figure 146: Compound 3.21 HMQC

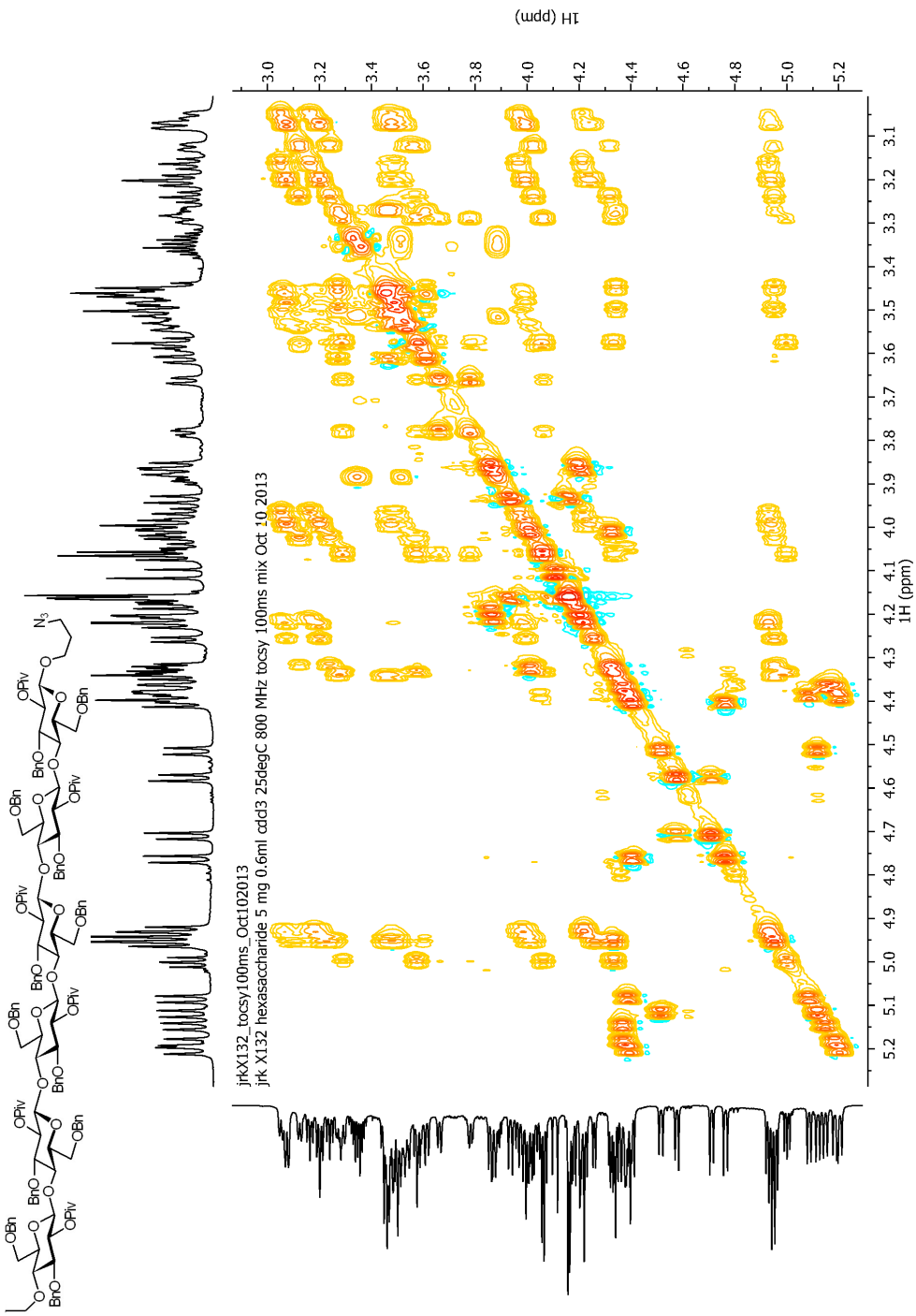


Figure 147: Compound 3.21 TOCSY

Appendix C

Thermodynamic models for CB[7]-methonium binding

Detailed thermodynamic models for chapter 4 were developed in collaboration with Dr. Yi Wang and are presented below.

Enthalpy

The binding enthalpy of the CB[7]•**1** complexes ($\Delta H^{CB[7]\bullet 1}$) can be separated into pairwise additive contributions from ligand-receptor interactions ($\Delta H_{\text{int}}^{CB[7]\bullet 1}$) and desolvation ($\Delta H_{desolv}^{CB[7]\bullet 1}$). $\Delta H_{\text{int}}^{CB[7]\bullet 1}$ can be written as

$$\Delta H_{\text{int}}^{CB[7]\bullet 1} = \Delta\langle U_1\rangle + \Delta\langle U_{CB[7]}\rangle + \Delta\langle U_{CB[7]-1}\rangle$$

Equation 52,

where $\Delta\langle\ldots\rangle = \langle\ldots\rangle_B - \langle\ldots\rangle_F$ and $\langle\ldots\rangle_B$ and $\langle\ldots\rangle_F$ denote the Boltzmann weighted averages of the bound and free states respectively, U_1 and $U_{CB[7]}$ are the internal energies of ligand **1** and CB[7] respectively, and $U_{CB[7]-1}$ is the interaction energy between CB[7] and **1** (not to be confused with the internal energy of the CB[7]•**1** complex). The host-guest interaction in the free state ($\langle U_{CB[7]-1}\rangle_F$) is set as the reference energy in our analysis. We further decompose $\Delta\langle U_1\rangle$ into the internal energy contributions from the Tris anchor ($\Delta\langle U_{Tris}\rangle$), the alkyl linker ($\Delta\langle U_{Link}\rangle$), and the change of Tris-linker interaction energy upon binding ($\Delta\langle U_{Tris-Link}\rangle$), which we assume to be unchanged across the ligand series in accordance with the anchor principle.[257, 258]

The change in CB[7] internal energy upon binding ($\Delta\langle U_{CB[7]}\rangle$) is assumed to contain two additive terms from Tris ($\Delta\langle U^{Tris}_{CB[7]}\rangle$) and the alkyl chain ($\Delta\langle U^{Link}_{CB[7]}\rangle$). Assuming the ligand-receptor interaction is pairwise additive, $\Delta H^{CB[7]\bullet 1}_{\text{int}}$ is:

$$\Delta H^{CB[7]\bullet 1}_{\text{int}} = \Delta\langle U_{Tris}\rangle + \Delta\langle U_{Link}\rangle + \Delta\langle U^{Link}_{CB[7]}\rangle + \langle U_{CB[7]-Tris}\rangle_B + \langle U_{CB[7]-Link}\rangle_B$$

Equation 53,

where $U_{CB[7]-Tris}$ and $U_{CB[7]-Link}$ are the CB[7]-Tris and CB[7]-linker interaction energies respectively. $\Delta H^{CB[7]\bullet 1}_{desolv}$ represents the desolvation enthalpy for the CB[7] cavity ($\Delta H_{desolv,CB[7]}$), the Tris anchor ($\Delta H_{desolv,Tris}$) and the alkyl chain ($\Delta H_{desolv,Link}$). The contribution per-methylene to binding enthalpy (ΔH_{CH_2}) in our system contains contributions from both desolvation and CB[7]-methylene interactions. Based on Schneider's linear free energy relationship,[259] we applied a linear approximation by partitioning $\Delta H^{CB[7]\bullet 1}_{b}$ measured via ITC as a function of the number of methylene groups in the linker (N):

$$\Delta H^{CB[7]\bullet 1}_{b} = N \bullet \Delta H_{CH_2} + \Delta H_0$$

Equation 54.

$N \bullet \Delta H_{CH_2}$ includes all of the terms in $\Delta H^{CB[7]\bullet 1}_{\text{int}}$ and $\Delta H^{CB[7]\bullet 1}_{desolv}$ pertaining to the alkyl linker:

$$N \bullet \Delta H_{CH_2} = \Delta\left\langle U_{Link} \right\rangle + \Delta\left\langle U_{CB[7]}^{Link} \right\rangle + \left\langle U_{CB[7]-Link} \right\rangle_B + \Delta H_{desolv,Link}$$

Equation 55;

ΔH_0 is defined to account for the remaining contributing terms in $\Delta H_b^{CB[7]\bullet 1}$.

Similarly, the binding enthalpy of the CB[7]•**2** complexes ($\Delta H_b^{CB[7]\bullet 2}$) is decomposed into an intrinsic interaction term ($\Delta H_{\text{int}}^{CB[7]\bullet 2}$) and a desolvation term ($\Delta H_{desolv}^{CB[7]\bullet 2}$), which can be expressed via the following expressions:

$$\begin{aligned}\Delta H_{\text{int}}^{CB[7]\bullet 2} = {} & \Delta\left\langle U_{Tris} \right\rangle + \Delta\left\langle U_{Link} \right\rangle + \Delta\left\langle U_{Am} \right\rangle + \Delta\left\langle U_{CB[7]}^{Tris} \right\rangle + \Delta\left\langle U_{CB[7]}^{Link} \right\rangle \\ & + \Delta\left\langle U_{CB[7]}^{Am} \right\rangle + \left\langle U_{CB[7]-Tris} \right\rangle_B + \left\langle U_{CB[7]-Link} \right\rangle_B + \left\langle U_{CB[7]-Am} \right\rangle_B\end{aligned}$$

Equation 56,

$$\Delta H_{desolv}^{CB[7]\bullet 2} = \Delta H_{desolv,CB[7]} + \Delta H_{desolv,Tris} + \Delta H_{desolv,Link} + \Delta H_{desolv,Am}$$

Equation 57,

where U_{Am}, $\Delta\left\langle U_{CB[7]}^{Am} \right\rangle$, $U_{CB[7]\bullet Am}$ and $\Delta H_{desolv,Am}$ are the internal energy of methonium, the change of CB[7] internal energy due to the binding of methonium, the CB[7]-methonium interaction energy and the enthalpy change for partial desolvation of methonium in the CB[7]•**2** complex. $\Delta H_b^{CB[7]\bullet 2}$ contains: 1) contributions whose magnitudes are constant throughout the series: ($\Delta\left\langle U_{Tris} \right\rangle + \Delta\left\langle U_{CB[7]}^{Tris} \right\rangle + \left\langle U_{CB[7]-Tris} \right\rangle_B + \Delta H_{desolv,CB[7]} + \Delta H_{desolv,Tris}$), because the Tris anchor binds at a constant position relative to the host; 2) contributions from the linker that can be approximated as $N \bullet \Delta H_{CH_2}$: ($\Delta\left\langle U_{Link} \right\rangle + \Delta\left\langle U_{CB[7]}^{Link} \right\rangle + \left\langle U_{CB[7]-Link} \right\rangle_B + \Delta H_{desolv,Link}$); and

3) contributions from methonium binding to CB[7]:

$(\Delta H_{b,\,Am} = \Delta\langle U_{Am}\rangle + \Delta\langle U^{Am}_{CB[7]}\rangle + \langle U_{CB[7]\text{–}Am}\rangle_B + \Delta H_{desolv,\,Am})$.

We next define the binding enthalpy difference ($\Delta\Delta H_b^{6\to N}$) between ligand **2** with N methylene groups in the linker ($\Delta H_b^{CB[7]\bullet 2}(N)$) and ligand **2d** with six methylene groups ($\Delta H_b^{CB[7]\bullet 2}(6)$), setting the CB[7]•**2d** complex as the reference state:

$$\Delta\Delta H_b^{6\to N} = (N-6)\Delta H_{CH_2} + \Delta\Delta^{6\to N}\langle U_{Am}\rangle + \Delta\Delta^{6\to N}\langle U^{Am}_{CB[7]}\rangle + \Delta^{6\to N}\langle U_{CB[7]\text{–}Am}\rangle_B + \Delta\Delta^{6\to N} H_{desolv,\,Am}$$

Equation 58,

where $\Delta\Delta^{6\to N}\langle ...\rangle$ is defined as $\Delta\langle ...\rangle(N) - \Delta\langle ...\rangle(6)$. Here, the last four terms of Equation 58 comprise the *net enthalpic effect* of methonium moving from its equilibrium position in the CB[7]•**2d** complex to another position closer to the CB[7] cavity:

$$\Delta\Delta H_{b,\,Am}^{6\to N} = \Delta\Delta H_b^{6\to N} - (N-6)\Delta H_{CH_2}$$

Equation 59.

$\Delta\Delta H_{b,\,Am}^{6\to N}$ contains both a desolvation term ($\Delta\Delta H_{desolv,\,Am}^{6\to N}$) and a term representing contributions from intrinsic interactions ($\Delta\Delta H_{\text{int},\,Am}^{6\to N}$; Equation 42):

$$\Delta\Delta H_{\text{int},\,Am}^{6\to N} = \Delta\Delta^{6\to N}\langle U_{Am}\rangle + \Delta\Delta^{6\to N}\langle U^{Am}_{CB[7]}\rangle + \Delta^{6\to N}\langle U_{CB[7]\text{–}Am}\rangle_B$$

Equation 60.

If we assume that the internal energy of methonium in the unbound state ($\langle U_{Am} \rangle_F$) is the same throughout the series, $\Delta\Delta^{6\to N}\langle U_{Am}\rangle$ then becomes:

$$\Delta\Delta^{6\to N}\langle U_{Am}\rangle = \Delta\langle U_{Am}\rangle(N) - \Delta\langle U_{Am}\rangle(6) = \langle U_{Am}\rangle_B(N) - \langle U_{Am}\rangle_B(6)$$

Equation 61.

By similarly setting the CB[7]-methonium interaction energy in the unbound state ($\langle U_{CB[7]\text{–}Am} \rangle_F$) to zero, $\Delta\Delta^{6\to N}\langle U_{CB[7]\text{–}Am}\rangle$ becomes:

$$\Delta\Delta^{6\to N}\langle U_{CB[7]\text{–}Am}\rangle = \langle U_{CB[7]\text{–}Am}\rangle_B(N) - \langle U_{CB[7]\text{–}Am}\rangle_B(6)$$

Equation 62.

Calculating the change in the CB[7] internal energy arising from repositioning methonium ($\Delta\Delta^{6\to N}\langle U^{Am}_{CB[7]}\rangle$) is not straightforward. Rather, only $\Delta\Delta^{6\to N}\langle U_{CB[7]}\rangle$ is computationally accessible. Replacing $\Delta\Delta^{6\to N}\langle U^{Am}_{CB[7]}\rangle$ with $\Delta\Delta^{6\to N}\langle U_{CB[7]}\rangle$, however, would lead to double counting of $\Delta\Delta^{6\to N}\langle U^{Link}_{CB[7]}\rangle$ and $\Delta\Delta^{6\to N}\langle U^{Tris}_{CB[7]}\rangle$, which are already accounted for. In light of the rigidity of the CB[7] structure, we assume $\Delta\Delta^{6\to N}\langle U^{Am}_{CB[7]}\rangle$ to be zero and $\Delta\Delta H^{6\to N}_{\text{int},\,Am}$ becomes:

$$\Delta\Delta H^{6\to N}_{\text{int},\,Am} = \langle U_{Am}\rangle_B(N) - \langle U_{Am}\rangle_B(6) + \langle U_{CB[7]\text{–}Am}\rangle_B(N) - \langle U_{CB[7]\text{–}Am}\rangle_B(6)$$

Equation 63,

where all components of $\Delta\Delta H^{6\to N}_{\text{int},\,Am}$ can be readily calculated from the ensemble averages of the internal energy of methonium in the bound state ($\langle U_{Am} \rangle_B$) and the CB[7]-

methonium interaction energy in the bound state ($\left\langle U_{CB[7]\text{–}Am} \right\rangle_B$) from trajectories of MD simulations of the CB[7]•**2** complexes. Therefore, by experimentally determining $\Delta\Delta H_{b,\,Am}^{6\rightarrow N}$ (Equation 41 or Equation 59), and calculating $\Delta\Delta H_{\text{int},\,Am}^{6\rightarrow N}$ via MD simulations (Equation 63), we obtain an estimate of $\Delta\Delta H_{desolv,\,Am}^{6\rightarrow N}$, the *net enthalpy change* for the incremental desolvation of methonium.

Heat capacity

The transfer of solute from water induces release of the solvation shell water to the bulk state;[295] the energetic signature for this process changes significantly as a function of temperature, for both ionic and neutral species.[296, 297] As such, the isobaric heat capacity changes during binding ($\Delta C_p = \frac{\partial H_b}{\partial T}$) are dominated by aqueous desolvation.[271, 272] Accordingly, we decompose the binding heat capacity of the CB[7]•**1** complexes ($\Delta C_p^{CB[7]\bullet 1}$) into the additive heat capacities for the CB[7] cavity ($\Delta C_p^{CB[7]}$) and for the Tris anchor (ΔC_p^{Tris}), which are independent of linker length (N), and the N-dependent heat capacity for the alkyl linker (ΔC_p^{Link}). A linear approximation of $\Delta C_p^{CB[7]\bullet 1}$ vs. N thus produces the per-methylene contribution to binding heat capacity ($\Delta C_p^{CH_2}$). Similar linear approximations can also be applied to the binding free energy ($\Delta G_b^{CB[7]\bullet 1}$) and entropy ($\Delta S_b^{CB[7]\bullet 1}$) to extract the corresponding per-methylene contribution (ΔG_{CH_2} and ΔS_{CH_2}; *vide infra*). For the CB[7]•**2** complexes

$$\Delta C_p^{CB[7]\bullet 2} = \Delta C_p^{CB[7]} + \Delta C_p^{Tris} + \Delta C_p^{Link} - \Delta C_p^{Am}$$

Equation 64.

Here, ΔC_p^{Am} is the contribution from methonium to the overall binding heat capacity ($\Delta C_p^{CB[7]\bullet 2}$). Similar to binding enthalpy, we use the CB[7]•**2d** complex as the reference state and define $\Delta\Delta C_p^{6\to N}$ as the difference in binding heat capacity for the CB[7]•**2d** and CB[7]•**2a-d** complexes. Correcting $\Delta\Delta C_p^{6\to N}$ for the change in linker contribution provides the *net change in heat capacity* due to the repositioning of methonium from CB[7]•**2d** to CB[7]•**2a-d** ($\Delta\Delta C_{p,\,Am}^{6\to N}$):

$$\Delta\Delta C_{p,\,Am}^{6\to N} = \Delta\Delta C_p^{6\to N} - (6-N)\Delta\Delta C_p^{CH_2}$$

Equation 65,

which largely reflects the degree of solvent reorganization upon the internalization of methonium, and is thus related to the temperature sensitivity of $\Delta\Delta H_{desolv,\,Am}^{6\to N}$.

Free energy and entropy

The approach similar to heat capacity can be applied to derive the net change in free energy and entropy upon the repositioning of methonium. The CB[7]•**1** complexes were used to quantify the free energy and entropy contributions of the methylene group (ΔG_{CH_2}, ΔS_{CH_2}) to the overall binding affinities. We partitioned binding free energy ($\Delta G_b^{CB[7]\bullet 1}$) and entropy ($\Delta S_b^{CB[7]\bullet 1}$) as functions of the number of methylene groups in the linker (N):

$$\Delta G_b^{CB[7]\bullet 1} = \Delta G_{CH_2}(N) + \Delta G_0$$

Equation 66,

$$\Delta S_b^{CB[7]\bullet 1} = \Delta S_{CH_2}(N) + \Delta S_0$$

Equation 67.

The linear regression produces an intercept terms ΔG_0 and ΔS_0, which comprise multiple phenomena including the interaction between the Tris anchor and CB[7] portal and the desolvation of CB[7] cavity.

Based on the assumption of simple group additivity, we treat ΔG_b and ΔS_b for the CB[7]•**2** complexes as:

$$\Delta G_b^{CB[7]\bullet 2} = \Delta G_{b,\,Am} + N\Delta G_{CH_2} + \Delta G_0'$$

Equation 68,

$$\Delta S_b^{CB[7]\bullet 2} = \Delta S_{b,\,Am} + N\Delta S_{CH_2} + \Delta S_0'$$

Equation 69,

where $\Delta G_{b,\,Am}$ and $\Delta S_{b,\,Am}$ are the free energy and entropy contributions from methonium group to the overall binding thermodynamics. By setting the CB[7]•**2d** complex as the reference state, and defining $\Delta\Delta G_b^{6\to N}$ and $\Delta\Delta S_b^{6\to N}$ as the difference in binding free energy and entropy for the formation of the CB[7]•**2d** and CB[7]•**2a-d** complexes, where N varies systematically between 3 and 5. $\Delta\Delta G_b^{6\to N}$ and $\Delta\Delta S_b^{6\to N}$ contain contributions from the change of linker length {$(N-6)\Delta G_{CH_2}$ and

$(N-6)\Delta S_{CH_2}$ } as well as a contribution associated with repositioning the methonium group from its equilibrium position in CB[7]•**2d** to the corresponding position in the CB[7]•**2a-c** complexes ($\Delta\Delta G_{b,\,Am}^{6\to N}$ and $\Delta\Delta S_{b,\,Am}^{6\to N}$), which are expressed as:

$$\Delta\Delta G_{b,\,Am}^{6\to N} = \Delta\Delta G_b^{6\to N} - (N-6)\Delta G_{CH_2} = \Delta\Delta G_b^{CB[7]\bullet 2a-c} + (N-6)\Delta G_{CH_2} - \Delta\Delta G_b^{CB[7]\bullet 2d}$$

Equation 70,

$$\Delta\Delta S_{b,\,Am}^{6\to N} = \Delta\Delta S_b^{CB[7]\bullet 2a-c} + (N-6)\Delta S_{CH_2} - \Delta\Delta S_b^{CB[7]\bullet 2d}$$

Equation 71.

Appendix D

Representative CB[7]•4.1b-e ITC titrations

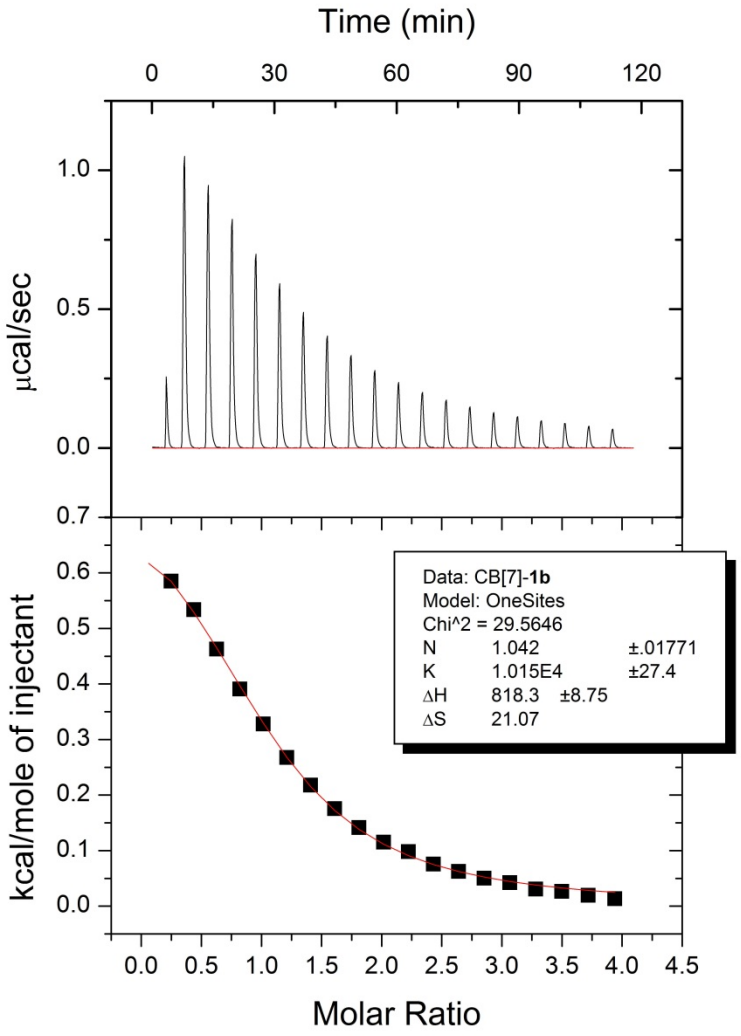


Figure 148: CB[7]•4.1b ITC titration at 303 K

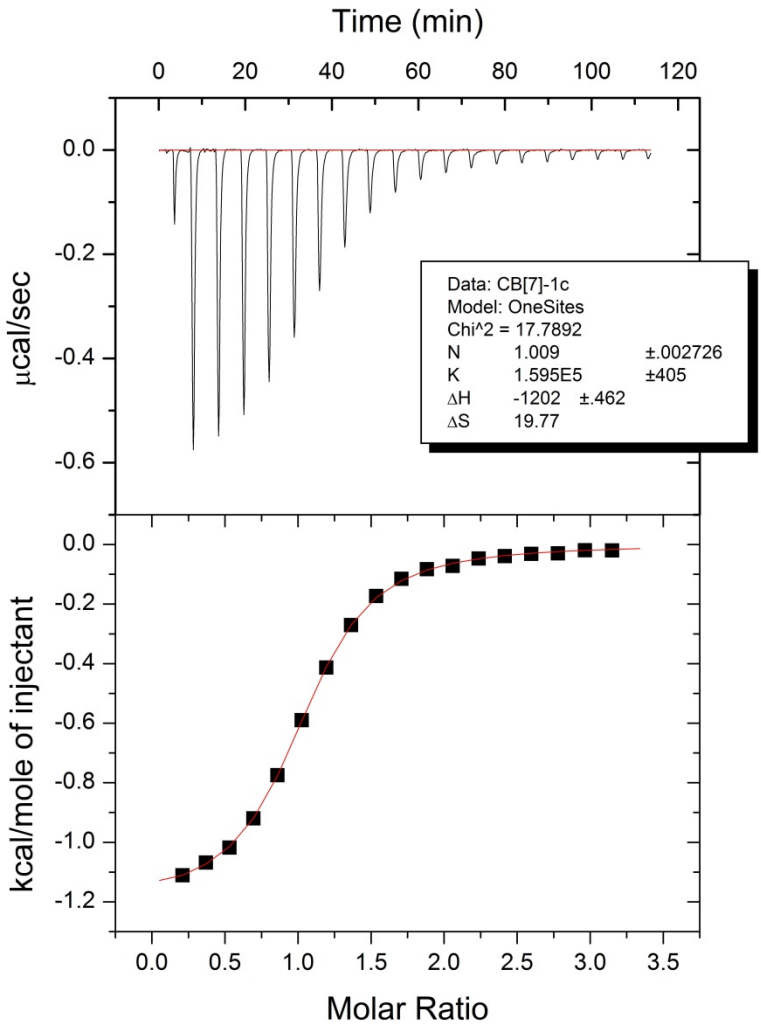


Figure 149: CB[7]•4.1c ITC titration at 298 K

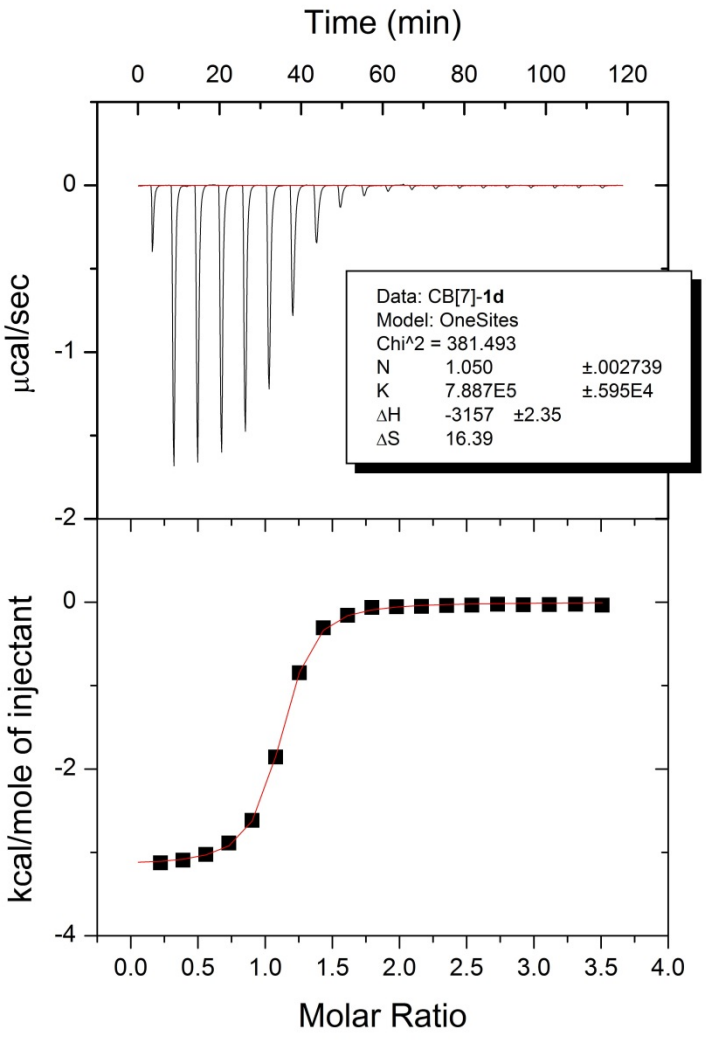


Figure 150: CB[7]•4.1d ITC titration at 298 K

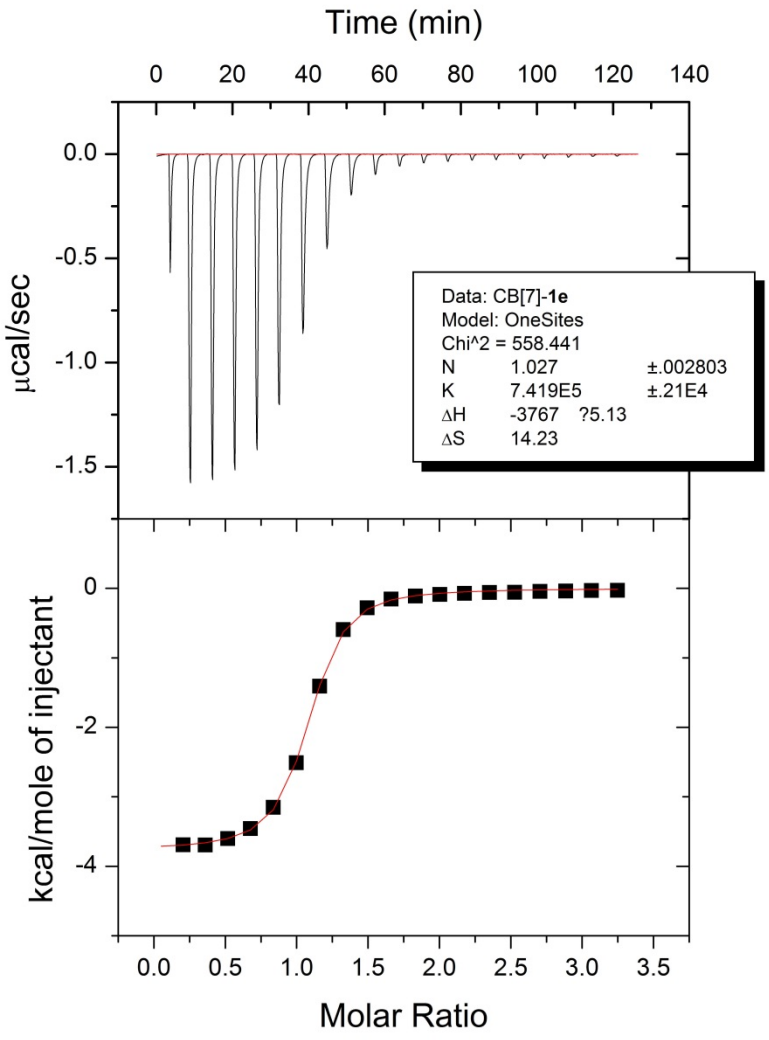


Figure 151: CB[7]•4.1e ITC titration at 298 K

Representative CB[7]•4.2a-d ITC titrations

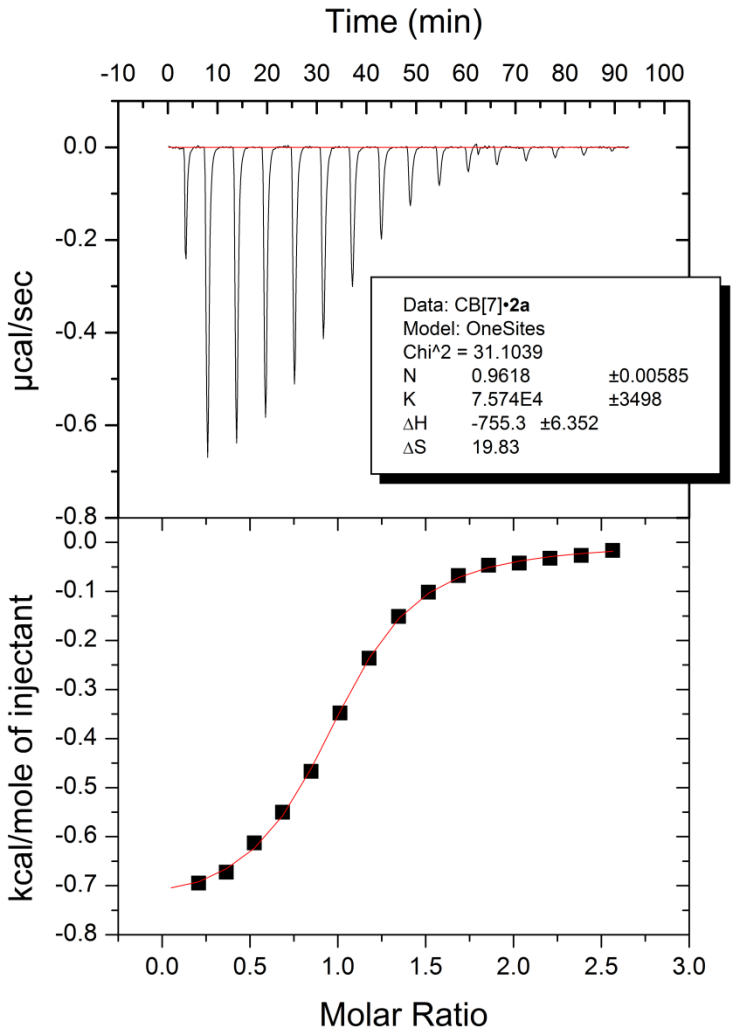


Figure 152: CB[7]•4.2a ITC titration at 298 K

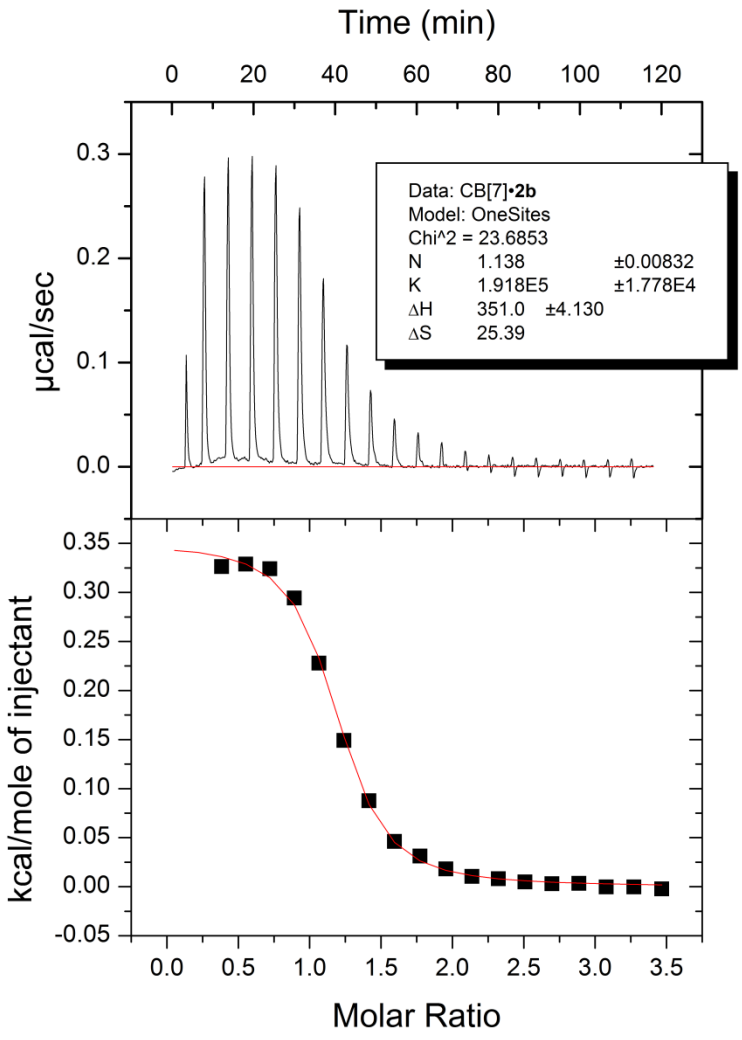


Figure 153: CB[7]•4.2b ITC titration at 303 K

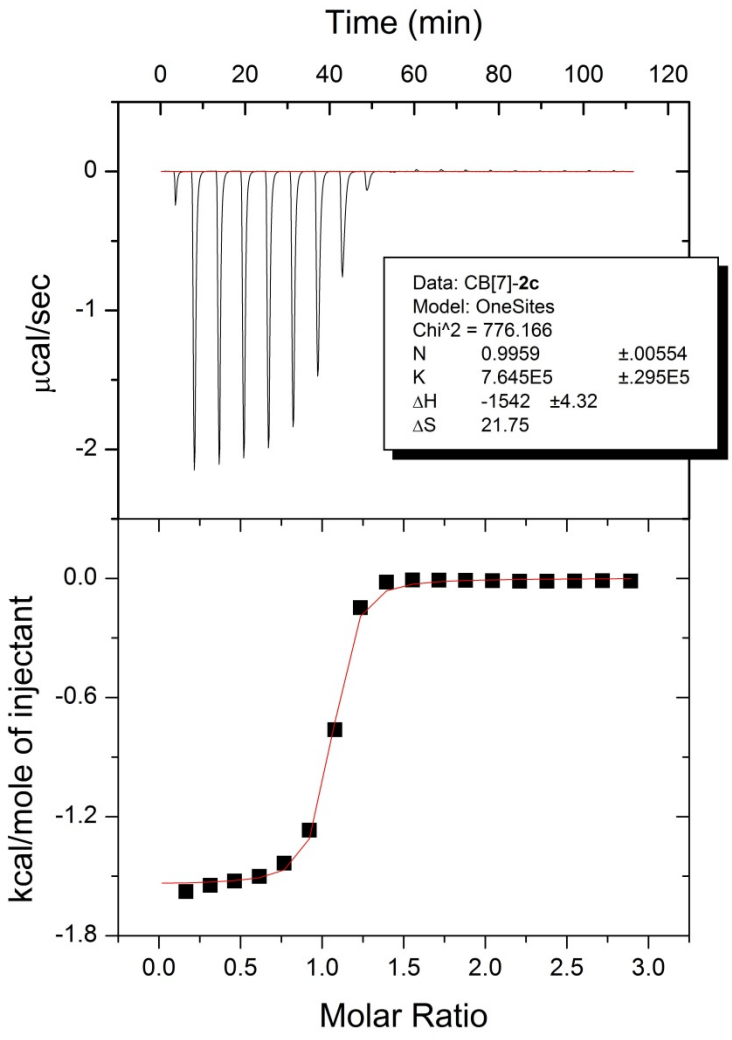


Figure 154: CB[7]•4.2c ITC titration at 298 K

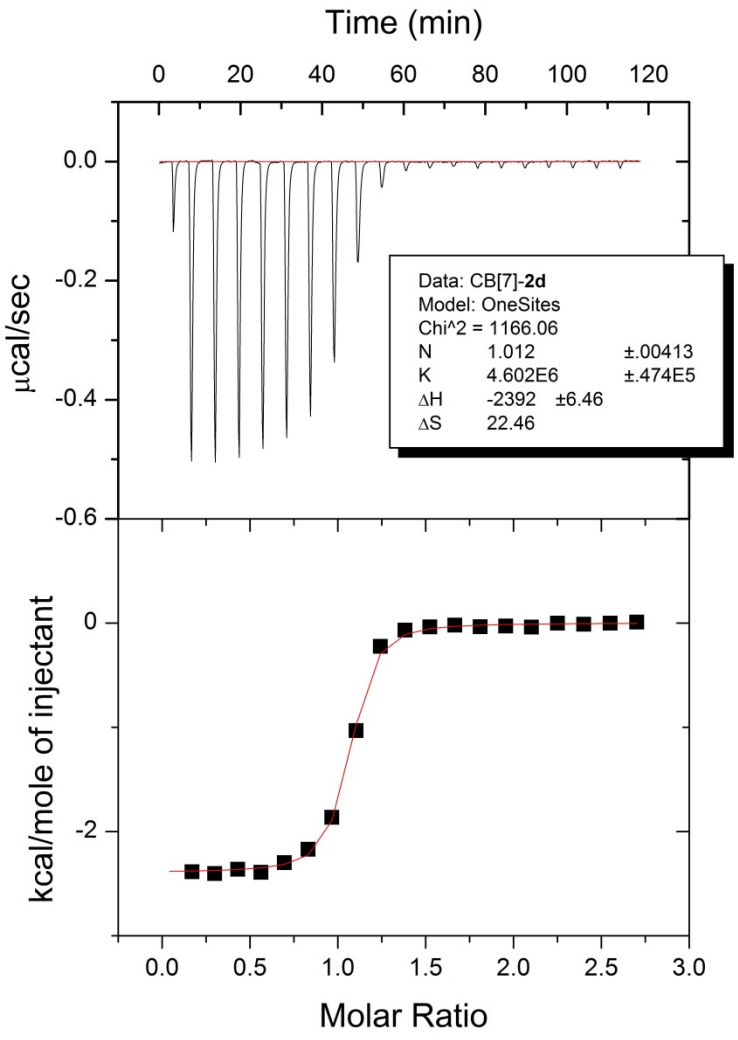


Figure 155: CB[7]•4.2d ITC titration at 298 K

References

1. Hammes, G. G. (1998) A golden era for understanding enzyme mechanisms, *Protein science : a publication of the Protein Society 7*, 799-802.

2. Gierasch, L. M., and Gershenson, A. (2009) Post-reductionist protein science, or putting Humpty Dumpty back together again, *Nature chemical biology 5*, 774-777.

3. Alberts, B. (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists, *Cell 92*, 291-294.

4. Bayer, E. A., Kenig, R., and Lamed, R. (1983) Adherence of Clostridium thermocellum to cellulose, *J Bacteriol 156*, 818-827.

5. Lamed, R., Setter, E., and Bayer, E. A. (1983) Characterization of a cellulose-binding, cellulase-containing complex in Clostridium thermocellum, *J Bacteriol 156*, 828-836.

6. Shoham, Y., Lamed, R., and Bayer, E. A. (1999) The cellulosome concept as an efficient microbial strategy for the degredation of insoluble polysaccharides, *Trends in Microbiology 7*, 275-281.

7. Petre, J., Longin, R., and Millet, J. (1981) Purification and properties of an endo-beta-1,4-glucanase from Clostridium thermocellum, *Biochimie 63*, 629-639.

8. Garcia-Martinez, D. V., Shinmyo, A., Madia, A., and Demain, A. (1980) Studies on Cellulase Production by Clostridium thermocellum, *European Journal of Applied Microbiology and Biotechnology 9*, 189-197.

9. Shinmyo, A., Garcia-Martinez, D. V., and Demain, A. L. (1979) Studies on the extracellular cellulolytic enzyme complex produced by clostridium thermocellum, *Journal of Applied Biochemistry 1*, 202-209.

10. Lamed, R., Setter, E., Kenig, R., and Bayer, E. A. (1983) Cellulosome: a discrete cell surface organelle of Clostridium thermocellum which exhibits separate antigenic, cellulose-binding and various cellulolytic activities, In *Biotechnology and Bioengineering Symposium*, pp 163-181.

11. Coughlan, M. P., Hon-Nami, K., Hon-Nami, H., Ljungdahl, L. G., Paulin, J. J., and Rigsby, W. E. (1985) The cellulolytic enzyme complex of Clostridium thermocellum is very large, *Biochemical and biophysical research communications 130*, 904-909.

12. Mayer, F., Coughlan, M. P., Mori, Y., and Ljungdahl, L. G. (1987) Macromolecular Organization of the Cellulolytic Enzyme Complex of Clostridium thermocellum as Revealed by Electron Microscopy, *Appl Environ Microbiol 53*, 2785-2792.

13. Garcia-Alvarez, B., Melero, R., Dias, F. M., Prates, J. A., Fontes, C. M., Smith, S. P., Romao, M. J., Carvalho, A. L., and Llorca, O. (2011) Molecular architecture and structural transitions of a Clostridium thermocellum mini-cellulosome, *J Mol Biol 407*, 571-580.

14. Kakiuchi, M., Isui, A., Suzuki, K., Fujino, T., Fujino, E., Kimura, T., Karita, S., Sakka, K., and Ohmiya, K. (1998) Cloning and DNA sequencing of the genes encoding Clostridium josui scaffolding protein CipA and cellulase CelD and identification of their gene products as major components of the cellulosome, *J Bacteriol 180*, 4303-4308.

15. Shoseyov, O., Takagi, M., Goldstein, M. A., and Doi, R. H. (1992) Primary sequence analysis of Clostridium cellulovorans cellulose binding protein A, *Proc. Natl. Acad. Sci. USA 89*, 3483-3487.

16. Shoseyov, O., Hamamoto, T., Foong, F., and Doi, R. H. (1990) Cloning of Clostridium cellulovorans endo-1,4-beta-glucanase genes, *Biochemical and biophysical research communications 169*, 667-672.

17. Shoseyov, O., and Doi, R. H. (1990) Essential 170-kDa subunit for degradation of crystalline cellulose by Clostridium cellulovorans cellulase, *Proc Natl Acad Sci U S A 87*, 2192-2195.

18. Doi, R. H., Kosugi, A., Muashima, K., Tamaru, Y., and San, S. O. (2003) Cellulosomes from mesophilic bacteria., *J. Bacteriol. 185*, 5907-5914.

19. Foong, F., Hamamoto, T., Shoseyov, O., and Doi, R. H. (1991) Nucleotide sequence and characteristics of endoglucanase gene engB from Clostridium cellulovorans, *Journal of general microbiology 137*, 1729-1736.

20. Shoseyov, O., Goldstein, M., Foong, F., Hamamoto, T., and Doi, R. H. (1991) Nucleotide sequence of Clostridium cellulovorans gene homologous to cyclic-AMP dependent kinase, *Nucleic acids research 19*, 1710.

21. Matano, Y., Park, J. S., Goldstein, M. A., and Doi, R. H. (1994) Cellulose promotes extracellular assembly of Clostridium cellulovorans cellulosomes, *J Bacteriol 176*, 6952-6956.

22. Doi, R. H., Goldstein, M., Hashida, S., Park, J. S., and Takagi, M. (1994) The Clostridium cellulovorans cellulosome, *Critical reviews in microbiology 20*, 87-93.

23. Abdou, L., Boileau, C., de Philip, P., Pages, S., Fierobe, H. P., and Tardif, C. (2008) Transcriptional regulation of the Clostridium cellulolyticum cip-cel operon: A complex mechanism involving a catabolite-responsive element, *Journal of Bacteriology 190*, 1499-1506.

24. Belaich, J. P., Tardif, C., Belaich, A., and Gaudin, C. (1997) The cellulolytic system of Clostridium cellulolyticum, *Journal of Biotechnology 57*, 3-14.

25. Blouzard, J. C., Bourgeois, C., de Philip, P., Valette, O., Belaich, A., Tardif, C., Belaich, J. P., and Pages, S. (2007) Enzyme diversity of the cellulolytic system produced by Clostridium cellulolyticum explored by two-dimensional analysis: Identification of seven genes encoding new dockerin-containing proteins, *Journal of Bacteriology 189*, 2300-2309.

26. Fierobe, H. P., Bagnaratardif, C., Gaudin, C., Guerlesquin, F., Sauve, P., Belaich, A., and Belaich, J. P. (1993) Purification and Characterization of Endoglucanase-C from Clostridium-Cellulolyticum - Catalytic Comparison with Endoglucanase-A, *European Journal of Biochemistry 217*, 557-565.

27. Fierobe, H. P., Pages, S., Belaich, A., Champ, S., Lexa, D., and Belaich, J. P. (1999) Cellulosome from Clostridium cellulolyticum: Molecular study of the dockerin/cohesin interaction, *Biochemistry-Us 38*, 12822-12832.

28. Bayer, E. A., Belaich, J.-P., Shoham, Y., and Lamed, R. (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides., *Ann. Rev. Microbiol. 58*, 521-554.

29. Giallo, J., Gaudin, C., Belaich, J. P., Petitdemange, E., and Cailletmangin, F. (1983) Metabolism of Glucose and Cellobiose by Cellulolytic Mesophilic Clostridium Sp Strain-H10, *Applied and Environmental Microbiology 45*, 843-849.

30. Lamed, R., Naimark, J., Morgenstern, E., and Bayer, E. A. (1987) Specialized cell surface structures in cellulolytic bacteria, *J Bacteriol 169*, 3792-3800.

31. Pohlschroder, M., Leschine, S. B., and Canale-Parola, E. (1994) Multicomplex cellulase-xylanase system of Clostridium papyrosolvens C7, *J Bacteriol 176*, 70-76.

32. Pohlschroder, M., Canale-Parola, E., and Leschine, S. B. (1995) Ultrastructural diversity of the cellulase complexes of Clostridium papyrosolvens C7, *J Bacteriol 177*, 6625-6629.

33. Leschine, S. B. (1995) Cellulose degradation in anaerobic environments, *Annual review of microbiology 49*, 399-426.

34. Nolling, J., Breton, G., Omelchenko, M. V., Makarova, K. S., Zeng, Q., Gibson, R., Lee, H. M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y. I., Tatusov, R. L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M. J., Bennett, G. N., Koonin, E. V., and Smith, D. R. (2001) Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum, *J Bacteriol 183*, 4823-4838.

35. Sabathe, F., and Soucaille, P. (2003) Characterization of the CipA scaffolding protein and in vivo production of a minicellulosome in Clostridium acetobutylicum, *J Bacteriol 185*, 1092-1096.

36. Sabathe, F., Belaich, A., and Soucaille, P. (2002) Characterization of the cellulolytic complex (cellulosome) of Clostridium acetobutylicum, *FEMS microbiology letters 217*, 15-22.

37. Ding, S. Y., Bayer, E. A., Steiner, D., Shoham, Y., and Lamed, R. (1999) A novel cellulosomal scaffoldin from Acetivibrio cellulolyticus that contains a family 9 glycosyl hydrolase, *J Bacteriol 181*, 6720-6729.

38. Ding, S. Y., Bayer, E. A., Steiner, D., Shoham, Y., and Lamed, R. (2000) A scaffoldin of the Bacteroides cellulosolvens cellulosome that contains 11 type II cohesins, *J Bacteriol 182*, 4915-4925.

39. Xu, Q., Bayer, E. A., Goldman, M., Kenig, R., Shoham, Y., and Lamed, R. (2004) Architecture of the Bacteroides cellulosolvens cellulosome: description of a cell surface-anchoring scaffoldin and a family 48 cellulase, *J Bacteriol 186*, 968-977.

40. Berger, E., Jones, W. A., Jones, D. T., and Woods, D. R. (1990) Sequencing and expression of a cellodextrinase (ced1) gene from Butyrivibrio fibrisolvens H17c cloned in Escherichia coli, *Molecular & general genetics : MGG 223*, 310-318.

41. Ding, S. Y., Rincon, M. T., Lamed, R., Martin, J. C., McCrae, S. I., Aurilia, V., Shoham, Y., Bayer, E. A., and Flint, H. J. (2001) Cellulosomal scaffoldin-like proteins from Ruminococcus flavefaciens, *J Bacteriol 183*, 1945-1953.

42. Kirby, J., Martin, J. C., Daniel, A. S., and Flint, H. J. (1997) Dockerin-like sequences in cellulases and xylanases from the rumen cellulolytic bacterium Ruminococcus flavefaciens, *FEMS microbiology letters 149*, 213-219.

43. Miron, J., Ben-Ghedalia, D., and Morrison, M. (2001) Invited review: adhesion mechanisms of rumen cellulolytic bacteria, *Journal of dairy science 84*, 1294-1309.

44. Ohara, H., Karita, S., Kimura, T., Sakka, K., and Ohmiya, K. (2000) Characterization of the cellulolytic complex (cellulosome) from Ruminococcus albus, *Bioscience, biotechnology, and biochemistry 64*, 254-260.

45. Ohara, H., Noguchi, J., Karita, S., Kimura, T., Sakka, K., and Ohmiya, K. (2000) Sequence of egV and properties of EgV, a Ruminococcus albus endoglucanase containing a dockerin domain, *Bioscience, biotechnology, and biochemistry 64*, 80-88.

46. Gerngross, U. T., Romaniec, M. P., Kobayashi, T., Huskisson, N. S., and Demain, A. L. (1993) Sequencing of a Clostridium thermocellum gene (cipA) encoding the cellulosomal SL-protein reveals an unusual degree of internal homology, *Molecular microbiology 8*, 325-334.

47. Bayer, E. A., Lamed, R., White, B. A., and Flint, H. J. (2008) From cellulosomes to cellulosomics, *Chem. Rec. 8*, 364-377.

48. Salamitou, S., Tokatlidis, K., Beguin, P., and Aubert, J. P. (1992) Involvement of separate domains of the cellulosomal protein S1 of Clostridium thermocellum in binding to cellulose and in anchoring of catalytic subunits to the cellulosome, *FEBS Lett 304*, 89-92.

49. Carvalho, A. L., Dias, F. M., Prates, J. A., Nagy, T., Gilbert, H. J., Davies, G. J., Ferreira, L. M., Romao, M. J., and Fontes, C. M. (2003) Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex, *Proc Natl Acad Sci U S A 100*, 13809-13814.

50. Karpol, A., Barak, Y., Lamed, R., Shoham, Y., and Bayer, E. A. (2008) Functional asymmetry in cohesin binding belies inherent symmetry of the dockerin module: insight into cellulosome assembly revealed by systematic mutagenesis, *Biochemical Journal 410*, 331-338.

51. Miras, I., Schaeffer, F., Beguin, P., and Alzari, P. M. (2002) Mapping by site-directed mutagenesis of the region responsible for cohesin-dockerin interaction on the surface of the seventh cohesin domain of Clostridium thermocellum CipA, *Biochemistry-Us 41*, 2115-2119.

52. Pinheiro, B. A., Proctor, M. R., Martinez-Fleites, C., Prates, J. A. M., Money, V. A., Davies, G. J., Bayer, E. A., Fontes, C. M. G. A., Fierobe, H. P., and Gilbert, H. J. (2008) The Clostridium cellulolyticum dockerin displays a dual binding mode for its cohesin partner, *J Biol Chem 283*, 18422-18430.

53. Schaeffer, F., Matuschek, M., Guglielmi, G., Miras, I., Alzari, P. M., and Beguin, P. (2002) Duplicated dockerin subdomains of Clostridium thermocellum endoglucanase CelD bind to a cohesin domain of the scaffolding protein CipA with distinct thermodynamic parameters and a negative cooperativity, *Biochemistry-Us 41*, 2106-2114.

54. Shimon, L. J., Bayer, E. A., Morag, E., Lamed, R., Yaron, S., Shoham, Y., and Frolow, F. (1997) A cohesin domain from Clostridium thermocellum: the crystal structure provides new insights into cellulosome assembly, *Structure 5*, 381-390.

55. Shimon, L. J., Frolow, F., Yaron, S., Bayer, E. A., Lamed, R., Morag, E., and Shoham, Y. (1997) Crystallization and preliminary X-ray analysis of a cohesin domain of the cellulosome from Clostridium thermocellum, *Acta crystallographica. Section D, Biological crystallography 53*, 114-115.

56. Valbuena, A., Oroz, J., Hervas, R., Vera, A. M., Rodriguez, D., Menendez, M., Sulkowska, J. I., Cieplak, M., and Carrion-Vazquez, M. (2009) On the remarkable mechanostability of scaffoldins and the mechanical clamp motif, *Proc Natl Acad Sci U S A 106*, 13791-13796.

57. Yaron, S., Morag, E., Bayer, E. A., Lamed, R., and Shoham, Y. (1995) Expression, purification and subunit-binding properties of cohesins 2 and 3 of the Clostridium thermocellum cellulosome, *FEBS Lett 360*, 121-124.

58. Pages, S., Belaich, A., Belaich, J. P., Morag, E., Lamed, R., Shoham, Y., and Bayer, E. A. (1997) Species-specificity of the cohesin-dockerin interaction between Clostridium thermocellum and Clostridium cellulolyticum: prediction of specificity determinants of the dockerin domain, *Proteins 29*, 517-527.

59. Lytle, B., Myers, C., Kruus, K., and Wu, J. H. (1996) Interactions of the CelS binding ligand with various receptor domains of the Clostridium thermocellum cellulosomal scaffolding protein, CipA, *J Bacteriol 178*, 1200-1203.

60. Poole, D. M., Morag, E., Lamed, R., Bayer, E. A., Hazlewood, G. P., and Gilbert, H. J. (1992) Identification of the cellulose-binding domain of the cellulosome subunit S1 from Clostridium thermocellum YS, *FEMS microbiology letters 78*, 181-186.

61. Bayer, E. A., Chanzy, H., Lamed, R., and Shoham, Y. (1998) Cellulose, cellulases and cellulosomes, *Curr. Opin. Struct. Biol. 8*, 548-557.

62. Beguin, P., and Alzari, P. M. (1998) The cellulosome of *Clostridium thermocellulum*., *Biochem. Soc. Trans. 26*, 178-185.

63. Beguin, P., and Lemaire, M. (1996) The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation., *Crit. Rev. Biochem. Mol. Biol. 31*, 201-236.

64. Fujino, T., Beguin, P., and Aubert, J. P. (1993) Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface, *J Bacteriol 175*, 1891-1899.

65. Leibovitz, E., and Beguin, P. (1996) A new type of cohesin domain that specifically binds the dockerin domain of the Clostridium thermocellum cellulosome-integrating protein CipA, *J Bacteriol 178*, 3077-3084.

66. Leibovitz, E., Ohayon, H., Gounon, P., and Beguin, P. (1997) Characterization and subcellular localization of the Clostridium thermocellum scaffoldin dockerin binding protein SdbA, *J Bacteriol 179*, 2519-2523.

67. Lemaire, M., Miras, I., Gounon, P., and Beguin, P. (1998) Identification of a region responsible for binding to the cell wall within the S-layer protein of Clostridium thermocellum, *Microbiology 144 (Pt 1)*, 211-217.

68. Smith, S. P., and Bayer, E. A. (2013) Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex, *Current opinion in structural biology 23*, 686-694.

69. Gefen, G., Anbar, M., Morag, E., Lamed, R., and Bayer, E. A. (2012) Enhanced cellulose degradation by targeted integration of a cohesin-fused beta-glucosidase into the Clostridium thermocellum cellulosome, *Proc Natl Acad Sci U S A 109*, 10298-10303.

70. Sukhumavasi, J., Ohmiya, K., Shimizu, S., and Ueno, K. (1988) Clostridium josui sp. nov., a cellulolytic, moderate thermophilic species from Thai compost, *International Journal of Systematic and Evolutionary Microbiology 38*, 179-182.

71. Pages, S., Belaich, A., Tardif, C., Reverbel-Leroy, C., Gaudin, C., and Belaich, J. P. (1996) Interaction between the endoglucanase CelA and the scaffolding protein CipC of the Clostridium cellulolyticum cellulosome, *J Bacteriol 178*, 2279-2286.

72. Doi, R. H. (2008) Cellulases of mesophilic microorganisms: cellulosome and noncellulosome producers, *Annals of the New York Academy of Sciences 1125*, 267-279.

73. Cornillot, E., Croux, C., and Soucaille, P. (1997) Physical and genetic map of the Clostridium acetobutylicum ATCC 824 chromosome, *J Bacteriol 179*, 7426-7434.

74. Rincon, M. T., Cepeljnik, T., Martin, J. C., Lamed, R., Barak, Y., Bayer, E. A., and Flint, H. J. (2005) Unconventional mode of attachment of the Ruminococcus flavefaciens cellulosome to the cell surface, *J Bacteriol 187*, 7569-7578.

75. Rincon, M. T., Martin, J. C., Aurilia, V., McCrae, S. I., Rucklidge, G. J., Reid, M. D., Bayer, E. A., Lamed, R., and Flint, H. J. (2004) ScaC, an adaptor protein carrying a novel cohesin that expands the dockerin-binding repertoire of the Ruminococcus flavefaciens 17 cellulosome, *J Bacteriol 186*, 2576-2585.

76. Rincon, M. T., Ding, S. Y., McCrae, S. I., Martin, J. C., Aurilia, V., Lamed, R., Shoham, Y., Bayer, E. A., and Flint, H. J. (2003) Novel organization and divergent dockerin specificities in the cellulosome system of Ruminococcus flavefaciens, *J Bacteriol 185*, 703-713.

77. Schaeffer, F., Matuschek, M., Guglielmi, G., Miras, I., Alzari, P. M., and Beguin, P. (2002) Duplicated dockerin subdomains of Clostridium thermocellum endoglucanase CelD bind to a cohesin domain of the scaffolding protein CipA with distinct thermodynamic parameters and a negative cooperativity, *Biochemistry-Us 41*, 2106-2114.

78. Mechaly, A., Fierobe, H. P., Belaich, A., Belaich, J. P., Lamed, R., Shoham, Y., and Bayer, E. A. (2001) Cohesin-dockerin interaction in cellulosome assembly: a single hydroxyl group of a dockerin domain distinguishes between nonrecognition and high affinity recognition, *J Biol Chem 276*, 9883-9888.

79. Mechaly, A., Yaron, S., Lamed, R., Fierobe, H. P., Belaich, A., Belaich, J. P., Shoham, Y., and Bayer, E. A. (2000) Cohesin-dockerin recognition in cellulosome assembly: experiment versus hypothesis, *Proteins 39*, 170-177.

80. Miras, I., Schaeffer, F., Beguin, P., and Alzari, P. M. (2002) Mapping by site-directed mutagenesis of the region responsible for cohesin-dockerin interaction on the

surface of the seventh cohesin domain of Clostridium thermocellum CipA, *Biochemistry-Us 41*, 2115-2119.

81. Bayer, E. A., Chanzy, H., Lamed, R., and Shoham, Y. (1998) Cellulose, cellulases and cellulosomes, *Current opinion in structural biology 8*, 548-557.

82. Beguin, P., and Aubert, J. P. (1994) The biological degradation of cellulose, *FEMS microbiology reviews 13*, 25-58.

83. Wada, M., Heux, L., and Sugiyama, J. (2004) Polymorphism of cellulose I family: reinvestigation of cellulose IVI, *Biomacromolecules 5*, 1385-1391.

84. Nishiyama, Y., Langan, P., and Chanzy, H. (2002) Crystal structure and hydrogen-bonding system in cellulose Ibeta from synchrotron X-ray and neutron fiber diffraction, *J Am Chem Soc 124*, 9074-9082.

85. Nishiyama, Y., Sugiyama, J., Chanzy, H., and Langan, P. (2003) Crystal structure and hydrogen bonding system in cellulose I(alpha) from synchrotron X-ray and neutron fiber diffraction, *J Am Chem Soc 125*, 14300-14306.

86. Ahola, S. (2008) Properties and interfacial behaviour of cellulose nanofibrils, In *Chemistry and Material Sciences*, Helsinki University of Technology, Espoo, Finland.

87. Cosgrove, D. J. (2005) Growth of the plant cell wall, *Nature reviews. Molecular cell biology 6*, 850-861.

88. Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S. (2002) Microbial cellulose utilization: fundamentals and biotechnology, *Microbiology and molecular biology reviews : MMBR 66*, 506-577, table of contents.

89. Demain, A. L., Newcomb, M., and Wu, J. H. (2005) Cellulase, clostridia, and ethanol, *Microbiology and molecular biology reviews : MMBR 69*, 124-154.

90. Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities, *The Biochemical journal 280 (Pt 2)*, 309-316.

91. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2013) The carbohydrate-active enzyme database (CAZy) in 2013, *Nucleic acids research*, 1-6.

92. Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H., 3rd, and Henrissat, B. (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5), *BMC evolutionary biology 12*, 186.

93. Bansal, P., Hall, M., Realff, M. J., Lee, J. H., and Bommarius, A. S. (2009) Modeling cellulase kinetics on lignocellulosic substrates, *Biotechnol Adv 27*, 833-848.

94. Irwin, D. C., Spezio, M., Walker, L. P., and Wilson, D. B. (1993) Activity Studies of 8 Purified Cellulases - Specificity, Synergism, and Binding Domain Effects, *Biotechnology and Bioengineering 42*, 1002-1013.

95. Jeoh, T., Ishizawa, C. I., Davis, M. F., Himmel, M. E., Adney, W. S., and Johnson, D. K. (2007) Cellulase digestibility of pretreated biomass is limited by cellulose accessibility, *Biotechnology and Bioengineering 98*, 112-122.

96. Woodward, J., Lima, M., and Lee, N. E. (1988) The role of cellulase concentration in determining the degree of synergism in the hydrolysis of microcrystalline cellulose, *The Biochemical journal 255*, 895-899.

97. Zverlov, V. V., Velikodvorskaya, G. A., and Schwarz, W. H. (2002) A newly described cellulosomal cellobiohydrolase, CelO, from Clostridium thermocellum: investigation of the exo-mode of hydrolysis, and binding capacity to crystalline cellulose, *Microbiology 148*, 247-255.

98. Lamed, R., Kenig, R., Morag, E., Calzada, J. F., Demicheo, F., and Bayer, E. A. (1991) Efficient cellulose solubilization by a combined cellulosome-beta-glucosidase system, *Applied Biochemistry and Biotechnology 21*, 173-183.

99. Chauvaux, S., Beguin, P., and Aubert, J. P. (1992) Site-directed mutagenesis of essential carboxylic residues in Clostridium thermocellum endoglucanase CelD, *J Biol Chem 267*, 4472-4478.

100. Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeyssens, M., Beguin, P., and Aubert, J. P. (1992) Crystal structure of a thermostable bacterial cellulose-degrading enzyme, *Nature 357*, 89-91.

101. White, A., Tull, D., Johns, K., Withers, S. G., and Rose, D. R. (1996) Crystallographic observation of a covalent catalytic intermediate in a beta-glycosidase, *Nature structural biology 3*, 149-154.

102. Kelly, J. A., Sielecki, A. R., Sykes, B. D., James, M. N., and Phillips, D. C. (1979) X-ray crystallography of the binding of the bacterial cell wall trisaccharide NAM-NAG-NAM to lysozyme, *Nature 282*, 875-878.

103. Koshland, D. E., Jr. (1953) Stereochemistry and the mechanism of enzymatic reactions, *Biological Reviews 28*, 416-436.

104. Gelb, M. H., Jain, M. K., Hanel, A. M., and Berg, O. G. (1995) Interfacial enzymology of glycerolipid hydrolases: lessons from secreted phospholipases A2, *Annu Rev Biochem 64*, 653-688.

105. Jain, M. K. B., O. G. (2002) *Interfacial Enzyme Kinetics*, John Wiley & Sons.

106. Parsiegla, G., Reverbel, C., Tardif, C., Driguez, H., and Haser, R. (2008) Structures of mutants of cellulase Ce148F of Clostridium cellulolyticum in complex with long hemithiocellooligosaccharides give rise to a new view of the substrate pathway during processive action, *Journal of Molecular Biology 375*, 499-510.

107. Lee, I., Evans, B. R., and Woodward, J. (2000) The mechanism of cellulase action on cotton fibers: evidence from atomic force microscopy, *Ultramicroscopy 82*, 213-221.

108. Uchihashi, T., Kodera, N., and Ando, T. (2012) Guide to video recording of structure dynamics and dynamic processes of proteins by high-speed atomic force microscopy, *Nature protocols 7*, 1193-1206.

109. Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Penttila, M., Ando, T., and Samejima, M. (2012) Visualization of cellobiohydrolase I from Trichoderma reesei moving on crystalline cellulose using high-speed atomic force microscopy, *Methods Enzymol 510*, 169-182.

110. Beldman, G., Voragen, A. G., Rombouts, F. M., Searle-van Leeuwen, M. F., and Pilnik, W. (1987) Adsorption and kinetic behavior of purified endoglucanases and exoglucanases from Trichoderma viride, *Biotechnol Bioeng 30*, 251-257.

111. Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G., and Haynes, C. A. (1996) Binding of the cellulose-binding domain of exoglucanase Cex from Cellulomonas fimi to insoluble microcrystalline cellulose is entropically driven, *Proc Natl Acad Sci U S A 93*, 12229-12234.

112. Kim, D. W., Jang, Y. H., and Jeong, Y. K. (1998) Adsorption kinetics and behaviour of two cellobiohydrolases from Trichoderma reesei on microcrystalline cellulose., *Biotechnology and Applied Biochemistry 27*, 97-102.

113. Cruys-Bagger, N., Elmerdahl, J., Praestgaard, E., Tatsumi, H., Spodsberg, N., Borch, K., and Westh, P. (2012) Pre-steady-state kinetics for hydrolysis of insoluble cellulose by cellobiohydrolase Cel7A, *J Biol Chem 287*, 18451-18458.

114. Bernardez, T. D., Lyford, K. A., and Lynd, L. R. (1994) Kinterics of the extracellular cellulases of Clostridium thermocellum on pretreated mixed hardwood and Avicel, *Appl. Microbiol. Biotechnol. 41*, 620-625.

115. Lee, S. B., Shin, H. S., Ryu, D. D., and Mandels, M. (1982) Adsorption of cellulase on cellulose: Effect of physicochemical properties of cellulose on adsorption and rate of hydrolysis, *Biotechnol Bioeng 24*, 2137-2153.

116. Steiner, W., Sattler, W., and Esterbauer, H. (1988) Adsorption of Trichoderma reesei cellulase on cellulose: experimental data and their analysis by different equations, *Biotechnol Bioeng 32*, 853-865.

117. Igarashi, K., Wada, M., Hori, R., and Samejima, M. (2006) Surface density of cellobiohydrolase on crystalline celluloses. A critical parameter to evaluate enzymatic kinetics at a solid-liquid interface, *FEBS J 273*, 2869-2878.

118. Zhang, Y. H., and Lynd, L. R. (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems, *Biotechnol Bioeng 88*, 797-824.

119. Bayer, E. A., Kenig, R., and Lamed, R. (1983) Adherence of Clostridium-Thermocellum to Cellulose, *Journal of Bacteriology 156*, 818-827.

120. Bayer, E. A., Skutelsky, E., Goldman, S., Rosenberg, E., and Gutnick, D. L. (1983) Immunochemical Identification of the Major Cell-Surface Agglutinogen of Acinetobacter-Calcoaceticus Rag-92, *Journal of general microbiology 129*, 1109-1119.

121. Fierobe, H.-P., Mechaly, A., Tardif, C., Belaich, A., Lamed, R., Shoham, Y., Belaich, J.-P., and Bayer, E. A. (2001) Design and production of active cellulosome chimeras: selective incorporation of dockerin-containing enzymes into defined functional complexes., *J. Biol. Chem. 276*, 21257-21261.

122. Fierobe, H.-P., Bayer, E. A., Tardif, C., Czyzek, M., Mechaly, A., Belaich, A., Lamed, R., Shoham, Y., and Belaich, J.-P. (2002) Degradation of cellulose substrates by cellulosome chimeras: substrate targeting versus proximity of enzyme components., *J. Biol. Chem. 277*, 49621-49630.

123. Mingardon, F., Chanal, A., Tardif, C., Bayer, E. A., and Fierobe, H.-P. (2007) Exploration of new geometries in cellulosome-like chimeras, *Appl. Environ. Microbiol. 73*, 7138-7149.

124. Ding, S.-Y., Xu, Q., Crowley, M., Zeng, Y., Nimlos, M., Lamed, R., Bayer, E. A., and Himmel, M. E. (2008) A biophysical perspective on the cellulosome: new opportunities for biomass conversion, *Curr. Opin. Biotechnol. 19*, 218-227.

125. el-Gogary, S., Leite, A., Crivellaro, O., Eveleigh, D. E., and el-Dorry, H. (1989) Mechanism by which cellulose triggers cellobiohydrolase I gene expression in Trichoderma reesei, *Proc Natl Acad Sci U S A 86*, 6138-6141.

126. Yu, A. H., and Saddler, J. N. (1995) Identification of essential cellulase components in the hydrolysis of steam-exploded birch substrate, *Biotechnology and Applied Biochemistry 21*, 185-202.

127. Morag, E., Bayer, E. A., and Lamed, R. (1992) Affinity digestion for the near-total recovery of purified cellulosome from Clostridium thermocellum, *Enzyme and Microbial Technology 14*, 289-292.

128. Ahsan, M., Matsumoto, M., Karita, S., Kimura, T., Sakka, K., and Ohmiya, K. (1997) Purification and characterization of the family J catalytic domain derived from the Clostridium thermocellum endoglucanase CelJ, *Bioscience, biotechnology, and biochemistry 61*, 427-431.

129. Ciruela, A., Gilbert, H. J., Ali, B. R., and Hazlewood, G. P. (1998) Synergistic interaction of the cellulosome integrating protein (CipA) from Clostridium thermocellum with a cellulosomal endoglucanase, *FEBS Lett 422*, 221-224.

130. Kataeva, I., Guglielmi, G., and Beguin, P. (1997) Interaction between Clostridium thermocellum endoglucanase CelD and polypeptides derived from the cellulosome-integrating protein CipA: stoichiometry and cellulolytic activity of the complexes, *The Biochemical journal 326 (Pt 2)*, 617-624.

131. Fierobe, H.-P., Mingardon, F., Mechaly, A., Belaich, A., Rincon, M. T., Pages, S., Lamed, R., Tardif, C., Belaich, J.-P., and Bayer, E. A. (2005) Action of designer celulosomes on homogenous versus complex substrates - controlled incorporation of three distinct enzymes intoa a defined trifuctional scaffoldin., *J. Biol. Chem. 280*, 16325-16334.

132. Vazana, Y., Marais, S., Barak, Y., Lamed, R., and Bayer, E. A. (2010) Interplay between Clostridium thermocellum family 48 and family 9 cellulases in

cellulosomal versus noncellulosomal states, *Appl. Environ. Microbiol. 76*, 3236-3243.

133. Tavares, G. A., Beguin, P., and Alzari, P. M. (1997) The crystal structure of a type I cohesin domain at 1.7 angstrom resolution, *Journal of Molecular Biology 273*, 701-713.

134. Spinelli, S., Fierobe, H. P., Belaich, A., Belaich, J. P., Henrissat, B., and Cambillau, C. (2000) Crystal structure of a cohesin module from Clostridium cellulolyticum: implications for dockerin recognition, *Journal of Molecular Biology 304*, 189-200.

135. Fontes, C. M., and Gilbert, H. J. (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates, *Annu Rev Biochem 79*, 655-681.

136. Adams, J. J., Pal, G., Yam, K., Spencer, H. L., Jia, Z., and Smith, S. P. (2005) Purification and crystallization of a trimodular complex comprising the type II cohesin-dockerin interaction of the cellulosome of Clostridium thermocellum, *Acta Crystallographica Section F-Structural Biology and Crystallization Communications 61*, 46-48.

137. Noach, I., Frolow, F., Jakoby, H., Rosenheck, S., Shimon, L. J., Lamed, R., and Bayer, E. A. (2005) Crystal structure of a type-II cohesin module from the Bacteriodes cellulosome reveals novel and distinctive secondary structural elements, *Journal of Molecular Biology 348*, 1-12.

138. Carvalho, A. L., Pires, V. M. R., Gloster, T. M., Turkenburg, J. P., Prates, J. A. M., Ferreira, L. M. A., Romao, M. J., Davies, G. J., Fontes, C. M. G. A., and Gilbert, H. J. (2005) Insights into the structural determinants of cohesin dockerin specificity revealed by the crystal structure of type II cohesin from Clostridium thermocellum SdbA, *Journal of Molecular Biology 349*, 909-915.

139. Noach, I., Frolow, F., Alber, O., Lamed, R., Shimon, L. J. W., and Bayer, E. A. (2009) Intermodular linker flexibility revealed from crystal structures of adjacent cellulosomal cohesins of Acetivibrio cellulolyticus, *Journal of Molecular Biology 391*, 86-97.

140. Bayer, E. A., Shimon, L. J., Shoham, Y., and Lamed, R. (1998) Cellulosomes-Structure and Ultrastructure, *Journal of Structural Biology 124*, 221-234.

141. Adams, J. J., Currie, M. A., Ali, S., Bayer, E. A., Jia, Z., and Smith, S. P. (2010) Insights into higher-order organization of the cellulosome revealed by a dissect-

and-build approach: crystal structure of interacting Clostridium thermocellum multimodular components, *Journal of Molecular Biology 396*, 833-839.

142. Adams, J. J., Pal, G., Jia, Z. C., and Smith, S. P. (2006) Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex, *Proc Natl Acad Sci U S A 103*, 305-310.

143. Hammel, M., Fierobe, H. P., Czjzek, M., Finet, S., and Receveur-Brechot, V. (2004) Structural insights into the mechanism of formation of cellulosomes probed by small angle X-ray scattering, *J Biol Chem 279*, 55985-55994.

144. Hammel, M., Fierobe, H. P., Czjzek, M., Kurkal, V., Smith, J. C., Bayer, E. A., Finet, S., and Receveur-Brechot, V. (2005) Structural basis of cellulosome efficiency explored by small angle X-ray scattering, *J Biol Chem 280*, 38562-38568.

145. Currie, M. A., Adams, J. J., Fauscher, F., Bayer, E. A., Jia, Z., and Smith, S. P. (2012) Scaffoldin conformation and dynamics revealed by a ternary complex from the clostridium thermocellum cellulosome, *J Biol Chem 287*, 26953-26961.

146. Currie, M. A., Cameron, K., Dias, F. M., Spencer, H. L., Bayer, E. A., Fontes, C. M., Smith, S. P., and Jia, Z. (2013) Small angle X-ray scattering analysis of Clostridium thermocellum cellulosome N-terminal complexes reveals a highly dynamic structure, *J Biol Chem 288*, 7978-7985.

147. Shimon, L. J. W., Pages, S., Belaich, A., Belaich, J. P., Bayer, E. A., Lamed, R., Shoham, Y., and Frolow, F. (2000) Structure of a family IIIa scaffoldin CBD from the cellulosome of Clostridium cellulolyticum at 2.2 angstrom resolution, *Acta Crystallogr D 56*, 1560-1568.

148. Lamed, R., Setter, E., and Bayer, E. A. (1983) Characterization of a Cellulose-Binding, Cellulase-Containing Complex in Clostridium-Thermocellum, *Journal of Bacteriology 156*, 828-836.

149. Johnson, P. E., Tomme, P., Joshi, M. D., and McIntosh, L. P. (1996) Interaction of soluble cellooligosaccharides with the N-terminal cellulose-binding domain of Cellulomonas fimi CenC 2. NMR and ultraviolet absorption spectroscopy, *Biochemistry 35*, 13895-13906.

150. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition, *Biochemical Journal 382*, 769-781.

151. Boraston, A. B. (2005) The interaction of carbohydrate-binding modules with insoluble non-crystalline cellulose is enthalpically driven, *Biochemical Journal 385*, 479-484.

152. Vazana, Y., Morais, S., Barak, Y., Lamed, R., and Bayer, E. A. (2010) Interplay between Clostridium thermocellum Family 48 and Family 9 Cellulases in Cellulosomal versus Noncellulosomal States, *Appl Environ Microb 76*, 3236-3243.

153. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition, *The Biochemical journal 382*, 769-781.

154. Shimon, L. J., Pages, S., Belaich, A., Belaich, J. P., Bayer, E. A., Lamed, R., Shoham, Y., and Frolow, F. (2000) Structure of a family IIIa scaffoldin CBD from the cellulosome of Clostridium cellulolyticum at 2.2 A resolution, *Acta crystallographica. Section D, Biological crystallography 56*, 1560-1568.

155. Torno, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., and Steitz, T. A. (1996) Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose, *The EMBO Journal 15*, 5739-5751.

156. Turon, X., Rojas, O. J., and Deinhammer, R. S. (2008) Enzymatic kinetics of cellulose hydrolysis: A QCM-D study, *Langmuir 24*, 3880-3887.

157. Arndt, P., Gerdes, R., Huschens, S., Pyplo-Schneiders, J., and Redlich, H. (2005) Preparation of cellulose oligomers from cellulose triacetate (standard procedure), *Cellulose 12*, 317–326.

158. Zhang, Y. H. P., and Lynd, L. R. (2003) Cellodextrin preparation by mixed-acid hydrolysis and chromatographic separation, *Analytical Biochemistry 322*, 225-232.

159. Wells, G. B., and Lester, R. L. (1979) Rapid Separation of Acetylated Oligosaccharides by Reverse-Phase High-Pressure Liquid-Chromatography, *Analytical Biochemistry 97*, 184-190.

160. El Rassi, Z. C., S. C. (1995) Chapter 2. Reversed-phase and hydrophobic interaction chromatography of carbohydrates and glycoconjugates; and Chapter 3. High performance interaction chromatography of carbohydrates with polar sorbents., *J. Chromatography Library 58*, 41-142.

161. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes, *Methods Enzymol 185*, 60-89.

162. Pages, S., Gal, L., Belaich, A., Gaudin, C., Tardif, C., and Belaich, J. P. (1997) Role of scaffolding protein CipC of Clostridium cellulolyticum in cellulose degradation, *J Bacteriol 179*, 2810-2816.

163. Charnock, S. J., Bolam, D. N., Nurizzo, D., Szabo, L., McKie, V. A., Gilbert, H. J., and Davies, G. J. (2002) Promiscuity in ligand-binding: The three-dimensional structure of a Piromyces carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose, *Proc Natl Acad Sci U S A 99*, 14077-14082.

164. Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter, *Anal Biochem 179*, 131-137.

165. Chervenak, M. C., and Toone, E. J. (1994) A Direct Measure of the Contribution of Solvent Reorganization to the Enthalpy of Ligand-Binding, *J Am Chem Soc 116*, 10533-10539.

166. Christenson, T., and Toone, E. J. (2003) Calorimetric evaluation of protein-carbohydrate affiniites, *Methods Enzymol 362*, 486-504.

167. Murphy, L., Baumann, M. J., Borch, K., Sweeney, M., and Westh, P. (2010) An enzymatic signal amplification system for calorimetric studies of cellobiohydrolases, *Anal Biochem 404*, 140-148.

168. Menguy, T., Chenevois, S., Guillain, F., le Maire, M., Falson, P., and Champeil, P. (1998) Ligand binding to macromolecules or micelles: Use of centrifugal ultrafiltration to measure low-affinity binding, *Analytical Biochemistry 264*, 141-148.

169. King, J. R., Bowers, C. M., and Toone, E. J. (2014) Specificy binding at the protein-cellulose interface observed with force spectroscopy, *Langmuir*, In preparation.

170. Hayashi, T., Yoshida, K., Park, Y. W., Konishi, T., and Baba, K. (2005) Cellulose metabolism in plants, *International review of cytology 247*, 1-34.

171. Linder, M., and Teeri, T. T. (1997) The roles and function of cellulose-binding domains, *Journal of Biotechnology 57*, 15-28.

172. Allen, S. G., Tanchak, O. M., Quirk, A., Raegen, A. N., Reiter, K., Whitney, R., Clarke, A. J., Lipkowski, J., and Dutcher, J. R. (2012) Surface plasmon resonance imaging of the enzymatic degredation of cellulose microfibrils, *Anal. Methods 4*, 3238-3245.

173. Mitsumori, M., Xu, L., Kajikawa, H., and Kurihara, M. (2002) Properties of cellulose-binding modules in endoglucanase F from Fibrobacter succinogenes S85 by means of surface plasmon resonance, *FEMS microbiology letters 214*, 277-281.

174. Abbott, D. W., and Boraston, A. B. (2012) Quantitative approaches to the analysis of carbohydrate-binding module function, *Methods Enzymol 510*, 211-231.

175. Snyder, P. W., Lee, G., Marszalek, P. E., Clark, R. L., and Toone, E. J. (2007) A stochastic, cantilever approach to the evaluation of solution phase thermodynamic quantities, *Proceedings of the National Academy of Sciences of the United States of America 104*, 2579-2584.

176. Zhang, M., Wu, S. C., Zhou, W., and Xu, B. (2012) Imaging and measuring single-molecule interaction between a carbohydrate-binding module and natural plant cell wall cellulose, *J Phys Chem B 116*, 9949-9956.

177. Zhang, M., Wang, B., and Xu, B. (2013) Measurements of single molecular affinity interactions between carbohydrate-binding modules and crystalline cellulose fibrils, *Physical chemistry chemical physics : PCCP 15*, 6508-6515.

178. Yokota, S., Matsuo, K., Kitaoka, T., and Wariishi, h. (2008) Specific interaction acting at a cellulose-binding domain/ cellulose interface for papermaking application, *BioResources 3*, 1030-1041.

179. Shestopalov, A. A., Clark, R. L., and Toone, E. J. (2010) Catalytic microcontact printing on chemically functionalized H-terminated silicon, *Langmuir 26*, 1449-1451.

180. Bowers, C. M., Carlson, D. A., Shestopalov, A. A., Clark, R. L., and Toone, E. J. (2012) A general and efficient cantilever functionalization technique for AFM molecular recognition studies, *Biopolymers 97*, 761-765.

181. Holmes-Farley, S. R., Reamey, R. H., McCarthey, T. J., Deutch, J., and Whitesieds, G. M. (1985) Acid-base behavior of carboxylic acid groups covalently attached at the surface of polyethylene: the usefulness of contact angle in following the ionization of surface functionality, *Langmuir 1*, 725-740.

182. Edgar, C. D., and Gray, D. G. (2003) Smooth model cellulose I surfaces from nanocrystal suspensions, *Cellulose 10*, 299-306.

183. Aulin, C., Ahola, S., Josefsson, P., Nishino, T., Hirose, Y., Osterberg, M., and Wagberg, L. (2009) Nanoscale cellulose films with different crystallinities and mesostructures--their surface properties and interaction with water, *Langmuir 25*, 7675-7685.

184. Bowers, C. M., Carlson, D. A., Rivera, M., Clark, R. L., and Toone, E. J. (2013) Effect of compressive force on unbinding specific protein-ligand complexes with force spectroscopy, *J Phys Chem B 117*, 4755-4762.

185. Lundquist, J. J., Kiburz, B. M., Wu, J. K., Gibbs, K. D., Jr., and Toone, E. J. (2002) Towards high affinity carbohydrate-binding proteins: Directed evolution of murine galectin-3, *Canadian Journal of Chemistry 80*, 999-1009.

186. Carrion-Vazquez, M., Oberhauser, A. F., Fowler, S. B., Marszalek, P. E., Broedel, S. E., Clarke, J., and Fernandez, J. M. (1999) Mechanical and chemical unfolding of a single protein: a comparison, *Proc Natl Acad Sci U S A 96*, 3694-3699.

187. Bogan, A. A., and Thorn, K. S. (1998) Anatomy of hot spots in protein interfaces, *J Mol Biol 280*, 1-9.

188. Clackson, T., and Wells, J. A. (1995) A hot spot of binding energy in a hormone-receptor interface, *Science 267*, 383-386.

189. Moretti, R., and Thorson, J. S. (2008) A comparison of sugar indicators enables a universal high-throughput sugar-1-phosphate nucleotidyltransferase assay, *Anal Biochem 377*, 251-258.

190. Park, J. T., and Johnson, M. J. (1949) A submicrodetermination of glucose, *J Biol Chem 181*, 149-151.

191. Zhang, Y. H., and Lynd, L. R. (2005) Determination of the number-average degree of polymerization of cellodextrins and cellulose with application to enzymatic hydrolysis, *Biomacromolecules 6*, 1510-1515.

192. Gupta, R., and Lee, Y. Y. (2008) Mechanism of cellulase reaction on pure cellulosic substrates, *Biotechnol Bioeng 102*, 1570-1581.

193. Chu, D., Deng, H., Zhang, X., Zhang, J., and Bao, J. (2012) A simplified filter paper assay method of cellulase enzymes based on HPLC analysis, *Appl Biochem Biotechnol 167*, 190-196.

194. Decker, S. R., Brunecky, R., Tucker, M. P., Himmel, M. E., and Selig, M. J. (2009) High-throughput screening techniques for biomass conversion, *Bioenergy Research 2*, 179-192.

195. Janis, J., Pulkkinen, P., Rouvinen, J., and Vainiotalo, P. (2007) Determination of steady-state kinetic parameters for a xylanase-catalyzed hydrolysis of neutral underivatized xylooligosaccharides by mass spectrometry, *Anal Biochem 365*, 165-173.

196. Bubner, P., Plank, H., and Nidetzky, B. (2013) Visualizing cellulase activity, *Biotechnol Bioeng 110*, 1529-1549.

197. Igarashi, K., Koivula, A., Wada, M., Kimura, S., Penttila, M., and Samejima, M. (2009) High speed atomic force microscopy visualizes processive movement of Trichoderma reesei cellobiohydrolase I on crystalline cellulose, *J Biol Chem 284*, 36186-36190.

198. Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Okamoto, T., Penttila, M., Ando, T., and Samejima, M. (2011) Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface, *Science 333*, 1279-1282.

199. Williams, B. A., and Toone, E. J. (1993) Calorimetric evaluation of enzyme kinetic parameters, *Journal of Organic Chemistry 58*, 3507-3510.

200. Constantinides, A., Myles, S. J., and Vieth, W. F. (1971) Glucono-delta-lactonase from Saccharomyces cerevisiae: extraction, purification and characterization, *Enzymologia biologica et clinica 43*, 121-128.

201. Chen, C. N., Chin, K. H., Wang, A. H., and Chou, S. H. (2008) The first crystal structure of gluconolactonase important in the glucose secondary metabolic pathways, *J Mol Biol 384*, 604-614.

202. Suzuki, K., Kawada, M., and Shimazono, N. (1961) Further studies on soluble lactonase, *The Journal of Biochemistry 49*, 448-449.

203. Gal, L., Pages, S., Gaudin, C., Belaich, A., Reverbel-Leroy, C., Tardif, C., and Belaich, J. P. (1997) Characterization of the cellulolytic complex (cellulosome) produced by Clostridium cellulolyticum, *Appl Environ Microbiol 63*, 903-909.

204. Reverbel-Leroy, C., Belaich, A., Bernadac, A., Gaudin, C., Belaich, J. P., and Tardif, C. (1996) Molecular study and overexpression of the Clostridium cellulolyticum celF cellulase gene in Escherichia coli, *Microbiology 142 (Pt 4)*, 1013-1023.

205. Reverbel-Leroy, C., Pages, S., Belaich, A., Belaich, J. P., and Tardif, C. (1997) The processive endocellulase CelF, a major component of the Clostridium cellulolyticum cellulosome: purification and characterization of the recombinant form, *J Bacteriol 179*, 46-52.

206. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures, *Protein expression and purification 41*, 207-234.

207. Armand, S., Drouillard, S., Schulein, M., Henrissat, B., and Driguez, H. (1997) A bifunctionalized fluorogenic tetrasaccharide as a substrate to study cellulases, *J Biol Chem 272*, 2709-2713.

208. Boyer, V., Fort, S., Frandsen, T. P., Schulein, M., Cottaz, S., and Driguez, H. (2002) Chemoenzymatic synthesis of a bifunctionalized cellohexaoside as a specific substrate for the sensitive assay of cellulase by fluorescence quenching, *Chemistry 8*, 1389-1394.

209. Francais, A., Urban, D., and Beau, J. M. (2007) Tandem catalysis for a one-pot regioselective protection of carbohydrates: the example of glucose, *Angewandte Chemie 46*, 8662-8665.

210. Kim, S. H., Augeri, D., Yang, D., and Kahne, D. (1994) Concise synthesis of the calicheamicin oligosaccharide using the sulfoxide glycosylation method, *J Am Chem Soc 116*, 1766-1775.

211. Abhishek, S., Abhijit, S., and Anup Kumar, M. (2011) Synthesis of thioglycosides in room temperature ionic liquid, *J Carb Chem 30*, 85-93.

212. Vedejs, E., and Daugulis, O. (1996) Dual activation in the esterification of hindered alcohols with anhydrides using MgBr2 and a tertiary amine, *The Journal of organic chemistry 61*, 5702-5703.

213. Wang, Y., King, J. R., Wu, P., Pelzman, D. L., Beratan, D. N., and Toone, E. J. (2013) Enthalpic signature of methonium desolvation revealed in a synthetic host-guest system based on cucurbit[7]uril, *J Am Chem Soc 135*, 6084-6091.

214. Wang, Y. (2012) Aqueous desolvation and molecular recognition: experimental and computational sutdies of a novel host-guest system based on cucurbit[7]uril, In *Department of Biochemistry*, Duke University, Durham, NC.

215. Bourne, Y., Taylor, P., and Marchot, P. (1995) Acetylcholinesterase Inhibition by Fasciculin - Crystal-Structure of the Complex, *Cell 83*, 503-512.

216. Sixma, T. K., and Smit, A. B. (2003) Acetylcholine binding protein (AChBP): A secreted glial protein that provides a high-resolution model for the extracellular domain of pentameric ligand-gated ion channels, *Annu. Rev. Biophys. Biomol. Struct. 32*, 311-334.

217. Celie, P. H. N., van Rossum-Fikkert, S. E., van Dijk, W. J., Brejc, K., Smit, A. B., and Sixma, T. K. (2004) Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures, *Neuron 41*, 907-914.

218. Pan, J., Chen, Q., Willenbring, D., Yoshida, K., Tillman, T., Kashlan, O. B., Cohen, A., Kong, X. P., Xu, Y., and Tang, P. (2012) Structure of the pentameric ligand-gated ion channel ELIC cocrystallized with its competitive antagonist acetylcholine, *Nat. Comm. 3*, 714.

219. Bourne, Y., Radic, Z., Sulzenbacher, G., Kim, E., Taylor, P., and Marchot, P. (2006) Substrate and product trafficking through the active center gorge of acetylcholinesterase analyzed by crystallography and equilibrium binding, *J. Biol. Chem. 281*, 29256-29267.

220. Dudev, T., and Lim, C. (2008) Metal binding affinity and selectivity in metalloproteins: Insights from computational studies, *Annu Rev Biophys 37*, 97-116.

221. Celio, M. R., Pauls, T. L., and Schwaller, B. (1996) *Guidebook to the calcium-binding proteins*, Sambrook & Tooze Publication at Oxford University Press, Oxford ; New York.

222. Malito, E., Sekulic, N., Too, W. C., Konrad, M., and Lavie, A. (2006) Elucidation of human choline kinase crystal structures in complex with the products ADP or phosphocholine, *J. Mol. Biol. 364*, 136-151.

223. Karlin, A. (2002) Emerging structure of the nicotinic acetylcholine receptors, *Nat. Rev. Neurosci. 3*, 102-114.

224. Duffy, E. M., Kowalczyk, P. J., and Jorgensen, W. L. (1993) Do Denaturants Interact with Aromatic-Hydrocarbons in Water, *J. Am. Chem. Soc. 115*, 9271-9275.

225. Gao, J., Chou, L. W., and Auerbach, A. (1993) The Nature of Cation-Pi-Binding - Interactions between Tetramethylammonium Ion and Benzene in Aqueous-Solution, *Biophys. J. 65*, 43-47.

226. Felder, C., Jiang, H. L., Zhu, W. L., Chen, K. X., Silman, I., Botti, S. A., and Sussman, J. L. (2001) Quantum/classical mechanical comparison of cation-pi interactions between tetramethylammonium and benzene, *J. Phys. Chem. A 105*, 1326-1333.

227. Meotner, M., and Deakyne, C. A. (1985) Unconventional Ionic Hydrogen-Bonds .1. Ch-Delta+...X - Complexes of Quaternary Ions with Normal-Donors and Pi-Donors, *J. Am. Chem. Soc. 107*, 469-474.

228. Marcus, Y. (1987) The Thermodynamics of Solvation of Ions .2. The Enthalpy of Hydration at 298.15-K, *J Chem Soc Farad T 1 83*, 339-349.

229. Marcus, Y. (1991) Thermodynamics of solvation of ions. Part 5.-Gibbs free energy of hydration at 298.15 K, *J. Chem. Soc., Faraday Trans. 87*, 2995-2999.

230. Roelens, S., and Torriti, R. (1998) Binding of acetylcholine and quaternary ammonium cations to macrocyclic and acyclic "phane" esters. Evaluation of the cation-pi primary interaction through adaptive aromatic hosts, *J. Am. Chem. Soc. 120*, 12443-12452.

231. Turner, J., Soper, A. K., and Finney, J. L. (1990) A Neutron-Diffraction Study of Tetramethylammonium Chloride in Aqueous-Solution, *Mol. Phys. 70*, 679-700.

232. Turner, J., Soper, A. K., and Finney, J. L. (1992) Water-Structure in Aqueous-Solutions of Tetramethylammonium Chloride, *Mol. Phys. 77*, 411-429.

233. Hulme, E. C., Soper, A. K., McLain, S. E., and Finney, J. L. (2006) The hydration of the neurotransmitter acetylcholine in aqueous solution, *Biophysical journal 91*, 2371-2380.

234. Parker, D., Kataky, R., Kelly, P. M., and Palmer, S. (1996) Selectivity in the binding and detection of charge diffuse ions, *Pure Appl. Chem. 68*, 1219-1223.

235. St-Jacques, A. D., Wyman, I. W., and Macartney, D. H. (2008) Encapsulation of charge-diffuse peralkylated onium cations in the cavity of cucurbit[7]uril, *Chem. Commun. (Cambridge, U. K.)*, 4936-4938.

236. Dvir, H., Silman, I., Harel, M., Rosenberry, T. L., and Sussman, J. L. (2010) Acetylcholinesterase: From 3D structure to function, *Chem.-Biol. Interact. 187*, 10-22.

237. Barthel, J., Krienke, H., and Kunz, W. (1998) *Physical chemistry of electrolyte solutions: modern aspects*, Steinkopf, Springer, Darmstadt, New York.

238. Maelicke, A., Fulpius, B. W., Klett, R. P., and Reich, E. (1977) Acetylcholine receptor. Responses to drug binding, *The Journal of biological chemistry 252*, 4811-4830.

239. Meyer, E. A., Castellano, R. K., and Diederich, F. (2003) Interactions with aromatic rings in chemical and biological recognition, *Angew. Chem. Int. Edit. 42*, 1210-1250.

240. Bissantz, C., Kuhn, B., and Stahl, M. (2010) A Medicinal Chemist's Guide to Molecular Interactions, *Journal of Medicinal Chemistry 53*, 5061-5084.

241. Moghaddam, S., Inoue, Y., and Gilson, M. K. (2009) Host-Guest Complexes with Protein-Ligand-like Affinities: Computational Analysis and Design, *J Am Chem Soc 131*, 4012-4021.

242. Rauwald, U., Biedermann, F., Deroo, S., Robinson, C. V., and Scherman, O. A. (2010) Correlating Solution Binding and ESI-MS Stabilities by Incorporating Solvation Effects in a Confined Cucurbit[8]uril System, *J. Phys. Chem. B 114*, 8606-8615.

243. Lagona, J., Mukhopadhyay, P., Chakrabarti, S., and Isaacs, L. (2005) The cucurbit[n]uril family, *Angewandte Chemie-International Edition 44*, 4844-4870.

244. Schneider, H. J. (2009) Binding Mechanisms in Supramolecular Complexes, *Angew. Chem. Int. Edit. 48*, 3924-3977.

245. Wyman, I. W., and Macartney, D. H. (2008) Cucurbit[7]uril host-guest complexes with small polar organic guests in aqueous solution, *Org. Biomol. Chem. 6*, 1796-1801.

246. Mock, W. L., and Shih, N. Y. (1986) Structure and Selectivity in Host Guest Complexes of Cucurbituril, *Journal of Organic Chemistry 51*, 4440-4446.

247. Nau, W. M., Florea, M., and Assaf, K. I. (2011) Deep Inside Cucurbiturils: Physical Properties and Volumes of their Inner Cavity Determine the Hydrophobic Driving Force for Host-Guest Complexation, *Israel Journal of Chemistry 51*, 559-577.

248. Biedermann, F., Uzunova, V. D., Scherman, O. A., Nau, W. M., and De Simone, A. (2012) Release of High-Energy Water as an Essential Driving Force for the High-Affinity Binding of Cucurbit[n]urils, *J. Am. Chem. Soc. 134*, 15318-15323.

249. Nguyen, C. N., Kurtzman Young, T., and Gilson, M. K. (2012) Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril, *J. Chem. Phys. 137*, 044101.

250. Masson, E., Ling, X. X., Joseph, R., Kyeremeh-Mensah, L., and Lu, X. Y. (2012) Cucurbituril chemistry: a tale of supramolecular success, *Rsc Adv 2*, 1213-1247.

251. Zhao, N., Liu, L., Biedermann, F., and Scherman, O. A. (2010) Binding Studies on CB[6] with a Series of 1-Alkyl-3-methylimidazolium Ionic Liquids in an Aqueous System, *Chem-Asian J 5*, 530-537.

252. Mukhopadhyay, P., Zavalij, P. Y., and Isaacs, L. (2006) High fidelity kinetic self-sorting in multi-component systems based on guests with multiple binding epitopes, *J. Am. Chem. Soc. 128*, 14093-14102.

253. Page, M. I., and Jencks, W. P. (1971) Entropic Contributions to Rate Accelerations in Enzymic and Intramolecular Reactions and Chelate Effect, *Proc. Natl. Acad. Sci. U. S. A. 68*, 1678-&.

254. Jencks, W. P. (1981) On the Attribution and Additivity of Binding-Energies, *P Natl Acad Sci-Biol 78*, 4046-4050.

255. Moghaddam, S., Yang, C., Rekharsky, M., Ko, Y. H., Kim, K., Inoue, Y., and Gilson, M. K. (2011) New Ultrahigh Affinity Host-Guest Complexes of Cucurbit[7]uril with Bicyclo[2.2.2]octane and Adamantane Guests: Thermodynamic Analysis and Evaluation of M2 Affinity Calculations, *J Am Chem Soc 133*, 3570-3581.

256. Rekharsky, M. V., Mori, T., Yang, C., Ko, Y. H., Selvapalam, N., Kim, H., Sobransingh, D., Kaifer, A. E., Liu, S. M., Isaacs, L., Chen, W., Moghaddam, S., Gilson, M. K., Kim, K. M., and Inoue, Y. (2007) A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy-entropy compensation, *P Natl Acad Sci USA 104*, 20737-20742.

257. Jencks, W. P. (1981) On the attribution and additivity of binding energies, *Proc Natl Acad Sci U S A 78*, 4046-4050.

258. Page, M. I., and Jencks, W. P. (1971) Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect, *Proc Natl Acad Sci U S A 68*, 1678-1683.

259. Schneider, H. J. (1994) Linear Free-Energy Relationships and Pairwise Interactions in Supramolecular Chemistry, *Chem. Soc. Rev. 23*, 227-234.

260. (we assume that the binding of ligands 1 and 2 to CB[7] results in complete dehydration of the CB[7] cavity).

261. Kim, K. (2002) Mechanically interlocked molecules incorporating cucurbituril and their supramolecular assemblies, *Chem. Soc. Rev. 31*, 96-107.

262. Wyman, I. W., and Macartney, D. H. (2010) Cucurbit[7]uril host-guest complexes of cholines and phosphonium cholines in aqueous solution, *Org. Biomol. Chem. 8*, 253-260.

263. Scarso, A., Trembleau, L., and Rebek, J. (2003) Encapsulation induces helical folding of alkanes, *Angew. Chem. Int. Edit. 42*, 5499-5502.

264. Mecozzi, S., and Rebek, J. (1998) The 55% solution: A formula for molecular recognition in the liquid state, *Chem-Eur J 4*, 1016-1022.

265. Plyasunov, A. V., and Shock, E. L. (2000) Thermodynamic functions of hydration of hydrocarbons at 298.15 K and 0.1 MPa, *Geochim. Cosmochim. Acta 64*, 439-468.

266. Plyasunov, A. V., and Shock, E. L. (2001) Group contribution values of the infinite dilution thermodynamic functions of hydration for aliphatic noncyclic hydrocarbons, alcohols, and ketones at 298.15 K and 0.1 MPa, *J. Chem. Eng. Data 46*, 1016-1019.

267. Kim, J., Jung, I. S., Kim, S. Y., Lee, E., Kang, J. K., Sakamoto, S., Yamaguchi, K., and Kim, K. (2000) New cucurbituril homologues: Syntheses, isolation, characterization, and X-ray crystal structures of cucurbit[n]uril (n=5, 7, and 8), *J. Am. Chem. Soc. 122*, 540-541.

268. MacCallum, J. L., Moghaddam, M. S., Chan, H. S., and Tieleman, D. P. (2007) Hydrophobic association of alpha-helices, steric dewetting, and enthalpic barriers to protein folding, *Proc. Natl. Acad. Sci. U. S. A. 104*, 6206-6210.

269. Baron, R., Setny, P., and McCammon, J. A. (2010) Water in Cavity-Ligand Recognition, *J Am Chem Soc 132*, 12091-12097.

270. Setny, P., Baron, R., and McCammon, J. A. (2010) How Can Hydrophobic Association Be Enthalpy Driven?, *Journal of Chemical Theory and Computation 6*, 2866-2871.

271. Syme, N. R., Dennis, C., Phillips, S. E. V., and Homans, S. W. (2007) Origin of heat capacity changes in a "Nonclassical" hydrophobic interaction, *ChemBioChem 8*, 1509-1511.

272. Prabhu, N. V., and Sharp, K. A. (2005) Heat capacity in proteins, *Annu. Rev. Phys. Chem. 56*, 521-548.

273. Shimizu, S., and Chan, H. S. (2001) Configuration-dependent heat capacity of pairwise hydrophobic interactions, *J. Am. Chem. Soc. 123*, 2083-2084.

274. Bourne, Y., Radic, Z., Sulzenbacher, G., Kim, E., Taylor, P., and Marchot, P. (2006) Substrate and product trafficking through the active center gorge of acetylcholinesterase analyzed by crystallography and equilibrium binding, *J. Biol. Chem. 281*, 29256-29267.

275. Rasaiah, J. C., Garde, S., and Hummer, G. (2008) Water in nonpolar confinement: From nanotubes to proteins and beyond, *Annual Review of Physical Chemistry 59*, 713-740.

276. Saenger, W. (1980) Cyclodextrin Inclusion-Compounds in Research and Industry, *Angewandte Chemie-International Edition in English 19*, 344-362.

277. Ashbaugh, H. S., and Pratt, L. R. (2006) Colloquium: Scaled particle theory and the length scales of hydrophobicity, *Rev Mod Phys 78*, 159-178.

278. Dill, K. A., Truskett, T. M., Vlachy, V., and Hribar-Lee, B. (2005) Modeling Water, the hydrophobic effect, and ion solvation., *Ann. Rev. Biophys. Biomol. Struct. 34*, 173-199.

279. Lemieux, R. U. (1996) How water provides the impetus for molecular recognition in aqueous solution, *Accounts of Chemical Research 29*, 373-380.

280. Syme, N. R., Dennis, C., Bronowska, A., Paesen, G. C., and Homans, S. W. (2010) Comparison of Entropic Contributions to Binding in a "Hydrophilic" versus "Hydrophobic" Ligand-Protein Interaction, *J Am Chem Soc 132*, 8682-8689.

281. Baldwin, R. L. (2012) Gas-liquid transfer data used to analyze hydrophobic hydration and find the nature of the Kauzmann-Tanford hydrophobic factor, *Proc. Natl. Acad. Sci. U. S. A. 109*, 7310-7313.

282. Yang, F., and Dearden, D. V. (2011) Gas Phase Cucurbit[n]uril Chemistry, *Isr. J. Chem. 51*, 551-558.

283. Armarego, W. L. F., and Chai, C. L. L. (2003) *Purification of Laboratory Chemicals*, Fifth ed., Elsevier, Burlington, MA, USA.

284. Sambrook, J., and Russell, D. W. (2001) Transformation of E. coli by Electroporation, In *Molecular Cloning: a laboratory manual* Third ed., pp 119-122, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

285. Sambrook, J., and Russell, D. W. (2001) The Inoue Method for Preparation and Transformation of Competent E. coli: "Ultra-competent" Cells, In *Molecular Cloning: a laboratory manual*, pp 112-115, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

286. Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins, *Biochemistry-Us 6*, 1948-1954.

287. Fierobe, H. P., Mechaly, A., Tardif, C., Belaich, A., Lamed, R., Shoham, Y., Belaich, J. P., and Bayer, E. A. (2001) Design and production of active cellulosome chimeras. Selective incorporation of dockerin-containing enzymes into defined functional complexes, *J Biol Chem 276*, 21257-21261.

288. Ludwig, F. R., and Jay, F. A. (1985) Reversible chemical cross-linking of the light-harvesting polypeptides of Rhodopseudomonas viridis, *European journal of biochemistry / FEBS 151*, 83-87.

289. Rivera, M., Morris, C., Carlson, D. A., Toone, E. J., Cole, D. G., and Clark, R. L. (2009) Minimizing tip-sample contact force in automated atomic force microscope based force spectroscopy, In *ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp 731-736, San Diego, CA, USA.

290. Rivera, M. (2008) Development of a State-of-the-art Atomic Force Microscope for Improved Force Spectroscopy, p 298, Duke University.

291. Miller, G. L. (1959) Use of dinitrosalycylic acid reagent for determination of reducing sugar, *Anal. Chem. 31*, 426-428.

292. Kartha, K. P. R., and Jennings, H. J. (1990) A simplified, one-pot preparation of acetobromo sugars from reducing sugars, *J Carb Chem 9*, 777-781.

293. Purves, C. B. (1929) Relations between rotary powar and structure in the sugar group. XXI. Beta-thiophenol glycosides of glucose, xylose, lactose, and cellobiose, *J. Am. Chem. Soc. 51*, 3619-3627.

294. Goggin, K. D., Hammen, P. D., Knutson, K. L., Lambert, J. F., Walinsky, S. W., and Watson, H. A., Jr. (1994) Commercial synthesis of alpha-D-cellobiosyl bromide heptaacetate, *Journal of chemical technology and biotechnology 60*, 253-256.

295. Bryantsev, V. S., Diallo, M. S., and Goddard, W. A. (2008) Calculation of solvation free energies of charged solutes using mixed cluster/continuum models, *J. Phys. Chem. B 112*, 9709-9719.

296. Dougherty, R. C. (1998) Temperature and pressure dependence of hydrogen bond strength: A perturbation molecular orbital approach, *J. Chem. Phys. 109*, 7372-7378.

297. Bennaim, A., and Marcus, Y. (1984) Solvation Thermodynamics of Nonionic Solutes, *J. Chem. Phys. 81*, 2016-2027.

Biography

Jason R. King

EDUCATION

Duke University, Durham, NC
PhD Degree in Chemistry **01/2014**
Advisor: Professor Eric J. Toone

University of Alabama at Birmingham, Birmingham, AL
BSc. in Chemistry **05/2008**
Advisor: Professor Sadanandan E. Velu

Alabama School of Mathematics and Science, Mobile, AL
Advanced Diploma **05/2004**

HONORS AND FELLOWSHIPS

Pelham Wilder teaching fellowship (2013); Duke Chemistry Department Recognition Award (November 2012); Burroughs Welcome organic chemistry fellow (2010-2011); Frank and Sarah McKnight prize in chemistry semi-finalist (2008); UAB Honors in Chemistry (2008); UAB Outstanding Undergraduate Student of Chemistry (2008); UAB Green and Gold Scholarship (2004-2008)

AFFILIATIONS

Phi Lambda Upsilon chemistry honors society (vice president 20010-2011; delegate to the national PLU congress of 2010; national PLU constitution and by-laws

committee 2010-2014; recruiter for PLU non-academic speaker 2009-

2011; PLU member 2008-present).

Duke Chemistry student assistant to the NMR staff (2009 – 2012)

TEACHING EXPERIENCE

Duke University, Durham, NC

Teaching Assistant:

Sophomore Organic Chemistry Recitation Instructor **5/2012-5/2013**

Sophomore Organic Chemistry Lab TA **9/2008- 5/2009, 9/2011-5/2012**

Senior Analytical Chemistry TA **9/2009- 12/2009**

PUBLICATIONS

Manuscripts

1. Wang, Y.; **King, J. R.**; Wu, P.; Pelzman, D. L.; Beratan, D. N.; Toone, E. J., Enthalpic signature of methonium desolvation revealed in a synthetic host-guest system based on curcurbit[7]uril. *Journal of the American Chemical Society*, **2013,** 135 (16), pp 6084–6091.

2. Chenna, B. C.; **King, J. R**.; Shinkre, B. A.; Glover, A. L.; Lucius, A. L.; Velu, S. E., Synthesis and structure activity relationship studies of novel Staphylococcus aureus Sortase A inhibitors. *European Journal of Medicinal Chemistry* **2010,** *45* (9), 3752-61.

3. Lee, Y. J.; **King, J. R.**; Chenna, B. C.; Owens, S. B.; Freeman, J. L.; Gray, G. M.; Velu, S. E., Synthesis and the Crystal Structure of (E)-2-(7-(3-(Thiophen-2-yl) acrylamido)-2,3-dihydro-5-oxobenzo[e][1,4]oxazepin-1(5H)-yl) ethyl acetate. *Journal of Chemical Crystallography* **2009,** *39* (12), 902-907.

4. Patel, S. P.; Nadkarni, D. H.; Murugesan, S.; **King, J. R.**; Velu, S. E., Azide-Mediated Detosylation of N-Tosylpyrroloiminoquinones and N-Tosylindole-4,7-quinones. *Synlett* **2008,** (18), 2864-2868.

5. Chenna, B. C.; Shinkre, B. A.; **King, J. R**.; Lucius, A. L.; Narayana, S. V.; Velu, S. E., Identification of novel inhibitors of bacterial surface enzyme Staphylococcus aureus Sortase A. *Bioorganic and Medicinal Chemistry Letters* **2008,** *18* (1), 380-5.

Manuscripts in preparation

6. **King, J. R**.; Bowers, C.M.; Toone, E. J., Specific binding at the cellulose binding module-cellulose interface observed by force spectroscopy. **In preparation**.

7. **King, J. R**. and Toone, E. J., An efficient synthesis of a fluorogenic cellohexaoside cellulase substrate using regioselective protection and chemoselective glycosyl activation strategies. **In preparation**.

Oral Presentations

Jason King, "In search of a biophysical understanding of the cellulosome." ARPA-E

Seminar, Washington, D.C., **September 2009**.

Poster Presentations

1. **King, J. R.**; Wang, Y.; Wu, P.; Pelzman, D. L.; Beratan, D. N.; Toone, E. J., "Enthalpic signature of methonium desolvation revealed in a synthetic host-guest system based on cucurbit[7]uril." Presented at the 246th annual meeting of the American Chemical Society, Indianapolis, IN, September 8-12, 2013; ORGN Paper 93 and SCIMIX Paper 36.

2. **King, J. R**.; Chenna, B. C.; Lucius, A. L.; Velu, S. E. "Discovery of Staphyloccocus aureus Sortase A inhibitors by in-silicon virtual screening." Presented at the 59th meeting of the Southeastern regional section of the American Chemical Society, Greenville, SC, October 24-27, 2007; Paper 366.